AFRL-HE-WP-TR-2004-0089

United States Air Force Research Laboratory

CONCEPT AND TECHNOLOGY EXPLORATION FOR TRANSPARENT HEARING

William L. Chapin
Agnieszka R. Jost
Bryan A. Cook
Fahri Surucu
Scott Foster

AuSIM, INC.
Mountain View, CA

Patrick M. Zurek
Joseph G. Desloge
Robert E. Beaudoin

SENSIMETRICS CORP.
Somerville, MA

Mark Bolas
Ian McDowall
Eric R. Lorimer

FAKESPACE LABORATORIES
Mountain View, CA

Barbara Shinn-Cunningham
Nathaniel Durlach

BOSTON UNIVERSITY
Boston, MA

May 2003

Interim Report for the Period November 2002 to May 2003

Human Effectiveness Directorate
Warfighter Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2004-0089

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

//Signed//

MARIS M. VIKMANIS
Chief, Warfighter Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*: May 2003

2. REPORT TYPE: Interim

3. DATES COVERED *(From - To)*: November 2002 to May 2003

4. TITLE AND SUBTITLE

Concept and Technology Exploration for Transparent Hearing

5a. CONTRACT NUMBER: F33657-97-D-6004

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER: 63231F

5d. PROJECT NUMBER: 2830

5e. TASK NUMBER: 30

5f. WORK UNIT NUMBER: 02

6. AUTHOR(S)

*William L. Chapin
Agnieszka R. Jost
Bryan A. Cook
Fahri Surucu
Scott Foster
**Patrick M. Zurek
Joseph G. Desloge
Robert E. Beaudoin
***Mark Bolas
Ian McDowall
Eric R. Lorimer

Cont'd in Block 13

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

*AuSIM, Inc.
Mountain View CA
**Sensimetrics Corp.
Somerville MA
***Fakespace Laboratories
Mountain View CA
****Boston University
Boston MA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory, Human Effectiveness Directorate
Warfighter Interface Division
Battlespace Acoustics Branch
Air Force Materiel Command
Wright Patterson AFB OH 45433-7901

10. SPONSOR/MONITOR'S ACRONYM(S): AFRL/HECB

11. SPONSOR/MONITOR'S REPORT NUMBER(S): AFRL-HE-WP-TR-2004-0089

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

****Barbara Shinn-Cunningham
Nathaniel Durlach

14. ABSTRACT

Modern militaries are challenged to physically protect open-field personnel from a great variety of life and effectiveness threats, including chemical, biological, laser, ballistic, and percussive weapons. Many chemical and biological threats require covering all orifices, including the ears, to achieve minimal protection. Additionally, warfighting involves operating in very close proximity to loud equipment, from which the noise can degrade an individual's auditory perception, and over time can degrade general performance. Common hearing protection and occlusion isolates the warfighter from the environment, deflating situational awareness, confidence, and effectiveness, thus putting the warfighter at high risk and compromising his ability to detect and assess threats. Often, soldiers are so uncomfortable with the isolation of hearing protection that they will choose to go without hearing protection and expose themselves to painful and harmful noise, which can result in deafness and reduced effectiveness as warfighters. This effort includes a survey of relevant head-borne hear-through auditory systems, a selection of approaches to a transparent hearing solution, implementation of the approaches, and evaluation.

15. SUBJECT TERMS

transparent hearing, directional audio, head-related transfer function, spatial localization, acoustic processing

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	140	David Darkow
					19b. TELEPHONE NUMBER *(Include area code)* 937.255.3660

This page intentionally left blank.

PREFACE

The work reported herein was conducted by AuSIM, Inc., Mountain View, CA, under Air Force contract F33657-97-D-6004, program element 63231F, work unit 28303002. The program was managed in the Battlespace Acoustics Branch, Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson AFB, OH. Dr. David Darkow was the technical monitor for the effort.

This page intentionally left blank.

Table of Contents

THIS PAGE IS INTENTIONALLY BLANK.

1 Introduction

1.1 This Document

This report completes the project entitled "Concept and Technology Exploration for Transparent Hearing Systems", funded by the US Air Force Research Laboratory at Wright-Patterson Air Force Base in collaboration with Natick Soldier Systems of the US Army. The document outlines the project as planned and details the project as executed. Given the importance and time criticality of determining a solution to the problem addressed, the project team exploited knowledge gained during the project, redirecting the plan as necessary to maximize exploration. This document outlines the goals of the project, provides an overview of previous relevant work, discusses the work planned for the project, details the work and its findings, and describes how a solution system could be integrated into a dismounted soldier's personal information system.

The intended audience for this document includes the project sponsors, the intermediate contract managers, designated reviewers, and future helmet system designers. Additionally, the report authors assume the document may be published to a wider audience. The designated reviewers may encompass professionals in the fields of hearing, signal processing, sensors, warfighting equipment, hearing enhancement/augmentation, and aural displays, who can give feedback and guidance to extensions of the project.

1.2 Motivation

Modern militaries are challenged to physically protect open-field personnel from a great variety of life and effectiveness threats, including chemical, biological, laser, ballistic, and percussive weapons. Many chemical and biological threats require covering all orifices, including the ears, to achieve minimal protection. Additionally, warfighting involves operating in very close proximity to loud equipment, from which the noise can degrade an individual's auditory perception, and over time can degrade general performance. Common hearing protection and occlusion isolates the warfighter from the environment, deflating situational awareness, confidence, and effectiveness, thus putting the warfighter at high risk and compromising his ability to detect and assess threats. Often, soldiers are so uncomfortable with the isolation of hearing protection that they will choose to go without hearing protection and expose themselves to painful and harmful noise, which can result in deafness and reduced effectiveness as warfighters.

Even without specific hearing protection, headgear in general distorts the normal presentation of sound to a human's ears, reducing the effectiveness of these omni-directional, spatially discriminating sensors. The challenge is to compensate for or minimize the negative acoustic effect of all headgear, and make hearing protection a positive outfitting for the warfighter.

1.3 Problem Statement

To form a problem statement, a broader view of the soldier's sense and use of hearing is here considered. In many types of warfare, the situational awareness of the dismounted soldier is severely degraded by an inability to hear and comprehend the acoustic environment. Three important classes of phenomena that contribute to this inability are described briefly in below.

Attenuation and alteration of acoustic inputs caused by headgear

Head-borne sensors and protective equipment, collectively called "headgear", is often employed to defend against threats and augment the soldiers' lethality and survivability. Often, this equipment covers the ears as well as other body parts. Even when the ears are not covered, this equipment's proximity to the head and shoulders can distort the incoming acoustic signals. Protective equipment intended to defend against acoustic threats, called hearing protection, causes severe attenuation as well as distortion. Distorted signals lose their identifying characteristics, including positional information.

Masking of some acoustic signals by other acoustic signals
In many cases, the environment contains acoustic signals that are so intense that the warfighter cannot hear other acoustic signals of importance (i.e., these other signals are "masked" by the intense signal). For example, a crucial verbal command may be completely masked by the sound of a tank, helicopter, machine gun, or explosion. Further, even if the verbal command is loud enough to be detected, the maskers may prevent the command from being understood.

Inability to sense acoustic signals because of temporary or permanent deafness
The overall level of acoustic energy in the warfighter's environment is often high enough to cause substantial temporary hearing loss or, in some cases, permanent deafness. Even without considering the acoustic effects resulting from enemy actions (bombs, specially designed acoustic weapons), the threat of deafness is severe. For example, shoulder-fired weapons can result in sounds of 180 dB SPL at the warfighter's ears. Such sounds, even though short in duration, can significantly degrade one's hearing abilities.

Through the examination of all aural influences about the soldier, solutions should be considered that not only deliver the acoustic environment with minimal distortion, but can also unmask positive signals, augment hearing in cases of loss, and provide aural cues that improve the warfighter's ability to localize and identify acoustic sources and to detect weak ones.

1.4 Project Objectives

The project to explore the issues relating to the above stated problem was designed to focus on four specific objectives:

1.4.1 Risk Reduction

The results of this proposed work should reduce technological risk for future related advanced technology programs by

1) narrowing solution space for related projects,
2) creating a body of knowledge for reference,
3) proving the viability of a solution for a previously unsolved problem,
4) providing guidelines for the design of a near optimal solution, and
5) estimating application development and implementation costs.

1.4.2 Metrics and Evaluation for Technologies and Solutions

Metrics, methods of evaluation, and evaluations of a representative set of Transparent Hearing solutions should provide immeasurable leverage for future related applications. A systems engineering approach should be taken, keeping an eye on integration with other system elements as well as to mechanical, processing, power, and weight constraints.

With well-developed metrics for Transparent Hearing evaluation, future application developers should be able to objectively compare approaches based on:

- Performance
- Energy Cost
- Manufacturing Cost
- Licensing Cost
- Maintenance Cost
- Reliability
- Effects of Headgear Accessories
- Potential for future enhancement
- Other factors

Evaluation of Transparent Hearing solutions with respect to both common tuning and optimal tuning to individual user characteristics should provide additional information for comparing these approaches.

1.4.3 Design Guidelines

General guidelines to designers of headgear are useful early in the design process. These guidelines should embody knowledge that applies to the full range of both military (dismounted, mounted, airborne, and at sea) and civilian applications (emergency and security personnel, industrial workers, etc.), wherever coordinated communication is required in environment possessing threats to health, life, and effectiveness.

1.4.4 Cost Estimation

The investigation of relative lifecycle costs, including development, manufacturing, maintenance, and replacement costs, and the identification of cost drivers for each approach should also provide valuable information for cost-effectiveness comparisons of the different approaches over the short and long term.

1.5 Method

To achieve the stated project objectives related to the stated problem, the project was designed to explore the concepts and technologies related to transparent hearing. The method of exploration includes:

- a survey of the existing knowledge base and product offerings,
- identification of the solution space in which all likely solutions may lie,
- selection of a representative sampling of possible solutions ("approaches") to for the basis of the exploration,
- implementation of the selected approaches to explore their characteristics in detail,
- definition of metrics for success in an approach towards a solution,
- evaluation of the approaches against the metrics, and
- a detailed report on the findings.

1.6 Terms

This section defined terms used in this document with a special emphasis for the context of this document.

- **Warfighter**
  For the context of this application, any dismounted personnel under threat and requiring situational awareness of the immediate surroundings to perform specific duties.
- **Headgear**
  All equipment worn on the head.
- **Pinna**
  Protruding appendage surrounding the ear canal providing direction-dependent resonances and partial obscuration for incoming sound signals. For the purpose of this document, pinna (*pl.* pinnae) includes the concha.
- **Concha**
  Largest cavity in the pinna providing prominent direction-dependent filtering characteristics. (*pl.* conchae)
- **Path**
  The trajectory of an acoustic wave from the emitter (sound source) to the receiver (listener).
- **Direct Path**
  The shortest trajectory of an acoustic wave between an emitter and a receiver.
- **Indirect Path**
  Acoustic wave signals that do not reach the receiver via the shortest path.
- **Controlled Path**
  Acoustic wave signals that are processed before reaching the receiver.
- **Occluded Hearing**
  A partially or fully obstructed direct path to the ear.
- **Protected Hearing**
  Hearing that has been shielded by passive or active devices with the use of which listeners will be able to maintain normal hearing capabilities subsequent to the occurrence of loud sounds that would ordinarily cause temporary or permanent hearing loss.
- **True Transparency**
  The inability to discriminate between unoccluded and occluded hearing.
- **Transparent Hearing**
  Perceptual restoration of hearing so the user can perform tasks equally well with and without hearing occlusion. The auditory tasks to be considered include signal detection in quiet and in noise, sound source localization, signal discrimination, signal identification, and speech intelligibility in noise.
- **Compensated Hearing**
  Hearing reinforcement that counterweighs a deficiency or impairment.
- **Natural Hearing Restoration**
  This term may be confused between Compensated Hearing and Transparent Hearing as described above and, therefore, will be avoided in this document.
- **Augmented Hearing**
  Hearing capability that is artificially boosted beyond natural hearing and may include hearing compensation, increased hearing sensitivity, augmented discrimination of signal from noise and aural-focusing on a particular direction or signal.

- **Automatic Gain Control (AGC)**
  In the context of this report, AGC refers to a controlled path that aims to adjust the gain such that the output signal remains below a threshold. Generically, the AGC term does not imply a particular method to achieve the attenuation (i.e. compression, limiter, clip).

1.7 System Objectives, Considerations, and Constraints

In addition to the psycho-acoustic properties described above, any transparent hearing system for the warfighter should be designed such that additional characteristics are considered or met.

1.7.1 Optimal Solution

An optimal solution would be a membrane or "force field" around the human head that

- is impervious to bullets and ballistic projectiles, protects against head injuries in falls such as paratroop jumps,
- is impervious to chemical poisons and biological germs and agents,
- is impervious to high-intensity optical energy such as lasers and bright flashes,
- is impervious to high-intensity aural energy above a specified level,
- permits normal oxygen, carbon-dioxide, and vapor transmission,
- permits normal optical transmission without distortion,
- permits normal aural transmission preserving sound wave structure across the spectrum,
- provides user-specific optical correction,
- provides user-specific aural correction,
- provides a means to display synthetic or electronically transmitted optical information,
- provides a means to display synthetic or electronically transmitted aural information,
- is comfortable to the user under all conditions, and
- requires very little energy.

It is not within the scope of this project to begin to achieve such a solution, but it is mentioned here so that sight of it is not lost in the focus on the components.

1.7.2 Transparent Hearing System Considerations

The need for transparency assumes the direct, uncontrolled path is obstructed by hearing protection, or more generally headgear. A system for achieving transparent hearing must necessarily replace the direct, uncontrolled path of sound-wave transmission from the acoustic environment to the warfighter's ears by an indirect, technologically-controlled, path. Elimination of the direct/uncontrolled path involves the use of passive and active signal–attenuation techniques. Achievement of the indirect/controlled path involves the use of microphones, earphones, and various forms of signal processing. Psychoacoustic elimination of the direct path is required not only for purposes of protection, but also for purposes of control. The task of achieving the desired controlled signals is greatly complicated by the addition of uncontrolled direct signals. The audio system for the controlled path must be realized in such a manner that it is compatible with the devices and procedures used to eliminate the direct/uncontrolled path.

Eliminating the direct/uncontrolled path is beyond the scope of this project. Therefore, it will be assumed that the direct/uncontrolled path is effectively eliminated, and the devices and procedures used to achieve this elimination will be ignored except for compatibility evaluation.

References to "Transparent Hearing System" mean a system that attenuates the direct, uncontrolled path to the point of psychoacoustic elimination and supplies an indirect, controlled path that supports transparent hearing.

1.7.3 Complete True Transparency

There are two reasons for which applicable transparent hearing solutions do not need to satisfy the constraint of true transparency or naturalness. First, satisfying such a criterion is arguably impossible. Second, such a system is not what is really needed. Understanding of this is already evident in the requirement that the system provide unnatural protection from acoustic trauma. Unnaturally good abilities to localize sound sources and to detect important signals in quiet and in background noise would, it may be assumed, also be appreciated. Basically, an audio system is needed that:

- provides acceptable performance,
- does not require a significant learning period,
- is robust,
- can be manufactured at low cost, and
- creates enthusiasm among potential users.

1.7.4 Compatibility Requirements

The Transparent Hearing System must be compatible with envisioned extensions or augmentations of the total warfighter audio system, as well as to headgear accessories.

1.7.4.1 Accessory Headgear

Transparent Hearing solutions should be designed and evaluated with respect to physical and acoustical compatibility with additional head-gear accessories such as laser detectors, night-vision, systems, various antennae, chem-bio masks, eye protection, comms systems, etc.

1.7.4.2 Advanced Auditory Displays

Transparent Hearing solutions should be designed and evaluated with respect to compatibility with advanced auditory displays, such as localized communications and aural information. For instance, Head-Related Transfer Functions for natural, transparent, and synthetic sounds should be compatible. The integration with data from multiple sensors including GPS, orientation, and night vision should be spatially coherent and intuitive. The leverage of Transparent Hearing sensors for other auditory displays should be considered, i.e. orientation sensing.

1.7.4.3 Advanced Augmented Hearing

Transparent Hearing solutions should be designed and evaluated with respect to compatibility with advanced augmented hearing solutions, such as supernormal listening, hearing-loss compensation, and remote battlefield sensing. Situation awareness can be increased beyond Transparent Hearing. The presentation of the surrounding aural environment may be completely controllable and even specifically augmented with user control. Techniques can provide augmented discrimination of signal from noise, or augmented aural-focusing on a particular direction or signal. The present objective is to provide transparent hearing with consideration for leveraging the same system for these augmented hearing techniques.

1.7.5 General-Applicability Requirements

The Transparent Hearing System must be able to fulfill its functions over a broad range of conditions. Dimensions along which conditions will vary include the acoustic environment, the paraphernalia worn by the warfighter, and the characteristics of the warfighter's auditory system. While the variability along these dimensions will require that the Transparent Hearing System be tunable, the ways and extent to which it must be tunable are currently uncertain.

1.7.6 Practicality Considerations

Practicality considerations include many items that will probably at some point become hard-specified constraints. They include:

- Energy Consumption
- Modularity and Interchangeability
- Field Serviceability
- Size, Weight, and Comfort
- Robustness and Ruggedness
- Costs: Energy, Manufacturing, Licensing, Maintenance, etc.

1.8 Non-Military Applicability

A good transparent hearing solution coupled with hearing protection promises applicability beyond the warfighter, deep into the private sector and civilian applications. All occupations that involve fairly noisy environments are candidates for hearing solutions derived from that described herein. Specific applicability varies from want to need with dependency on situational awareness and the inherent health risk.

- industrial equipment operators
- urban firefighters
- aviation ground crews
- outdoor sportsmen
- football coaches
- construction workers
- wildfire firefighters
- broadcast crews
- event security
- auto-racing teams

This broad applicability means that a solution could save lives, improve productivity, and reduce health risks for hundreds of thousands of everyday people, not just the elite warfighter on a rare high-risk mission.

THIS PAGE IS INTENTIONALLY BLANK.

2 Background

This section contains background information relevant to development of the Transparent Hearing System. It includes material on normal hearing performance, on alterations in hearing due to changes in biological processing (e.g. hearing loss) or to changes in the input signals (e.g. head-worn gear disturbance), and on adaptation to such alterations.

2.1 Normal Hearing

2.1.1 General Perceptual Characteristics

In general, a listener's objective performance can be characterized by two parameters: resolution and response bias. Resolution measures the extent to which the listener can discriminate slightly different stimuli. Response bias measures the extent to which the listener tends to make one response over another, independent of the stimulus. Significant biases can generally be eliminated by short training periods in which correct-response feedback is presented to the listener. Poor resolution, on the other hand, tends to reflect fundamental limitations in auditory processing and is less susceptible to improvement by training.

2.1.1.1 Just Noticeable Differences (JND)

Results of psychoacoustic studies [40][93] indicate that:

- the detection threshold for sounds in a completely quiet background (the "absolute threshold") is of the order of 0 dB SPL in the mid-frequencies;
- the JND in frequency is of the order of 3% of the reference frequency;
- the JND in level is of the order of 1 dB; and
- the JND in duration is of the order of 10% of the reference duration [99]

Specific to spatial localization, studies show [95] that:

- the JND in source azimuth near the frontal position is 2 or 3 degrees;
- the JND in source azimuth near the interaural axis is on the order of 20 degrees, and
- the JND in source elevation is on the order of 20 degrees; and
- the JND in source distance is relatively poor unless one has excellent a priori information on the signal's intensity level at the source[35][93].

The JND for localization angle irrespective of axis is more formally referred to as Minimum Audible Angle (MAA).

It is important to note, however, that these resolution figures represent the results obtained under ideal laboratory conditions. They are likely to be substantially degraded by the presence of competing sounds that tend to mask the "target" signal, of echoes and reverberation in the acoustic environment, and of uncertainty in the acoustic stimuli.

2.1.1.2 Interference

Monaural masked thresholds are roughly equal to the value or power of signal cues required to achieve a signal-to-noise ratio of unity at the output of the relevant critical bands[1]. Binaural masked thresholds, including both the "better ear effect" and the results of binaural interaction, are often 10-20 dB lower than the measured thresholds for a single ear [35]. Obviously, as the target-signal level approaches its masked threshold, discrimination and recognition performance, as well as detection performance, are degraded.

Spatial localization tends to be degraded by the presence of echoes and reverberation in the environment; however, the amount of degradation is limited to a certain extent by the "precedence effect", whereby the impression of location is dominated by the interaural cues carried by the direct acoustic wave [60][149].

[1] Critical bands are the psycho-acoustically-determined auditory filters present in the biological processing.

2.1.1.3 Stimulus and Identification

Performance also tends to be degraded when the listener is uncertain about the sounds to be presented or the choices to be made in response to the received sounds. Most of the data obtained in the laboratory are obtained under conditions where uncertainty of these types is minimized. Although the effects of uncertainty have been studied by a few individuals for a number of years [137], it is only recently that this area has become a central focus of psychoacoustic research.

Finally, it should be noted that for a set of sounds in which the members differ by only one or two dimensions, identification performance (which is strongly limited by memory constraints) is much worse than would be expected on the basis of discrimination results. For example, it is impossible to reliably identify the intensity of a sound when the number of intensities in the set exceeds roughly 7, even when the intensities are separated by many JND's [39].

2.1.2 Spatial Localization

Spatial localization refers to the ability of human listeners to judge the direction and distance of environmental sound sources. To determine the direction of a sound, the auditory system relies on various physical cues. Sound waves emanating from a source travel in all directions away from the source. Some waves travel to the listener using the most direct path (direct sound), while others reflect off of walls and objects before reaching the listener's ears (indirect sound). The direct sound carries information about the location of the source relative to the listener. Indirect sound informs the listener about the space, and the relation of the source location to that space.

2.1.2.1 Interaural Differences

Because of the ears' spatial disparity and the mass between them, they each receive a different version of the arriving sound. The ear that is closest to the sound (ipsilateral ear) will receive the sound earlier and at a greater intensity or level than the ear farther away from the source (contralateral ear). The differences in time of arrival and in level are referred to as the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD)[2] respectively.

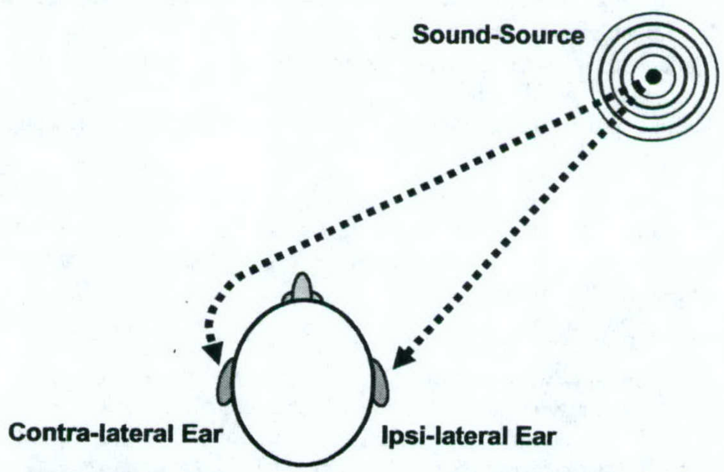


Figure 1: Schematic showing ipsi-lateral ear (near) versus the contra-lateral ear (far). The signal arrives at the contra-lateral ear later, attenuated, and shadowed in the high-frequencies (above 1 kHz) as compared to the ipsi-lateral.

[2] Interaural Level Difference (ILD) is also referred to as Interaural Intensity Difference (IID).

To a good approximation, ITD is independent of frequency. However, for narrowband signals, the auditory system is incapable of sensing the ITD much above 1500 Hz due to phase ambiguities. If, on the other hand, the signal is sufficiently broadband (so that the phase ambiguities can be resolved), then ITD can be sensed at high frequencies as well as low frequencies (although with somewhat less accuracy). At low frequencies, and for a reference ITD of 0 μsec, the ITD JND is roughly 10 μsec [95].

Unlike ITD, the interaural parameter ILD depends strongly on frequency, decreasing more or less monotonically in magnitude as frequency is lowered, because the head-shadow effect diminishes as the wavelength of the sound becomes appreciable relative to the size of the head. Thus, even though the auditory system maintains an interaural level JND of roughly 1 dB at all frequencies for a reference ILD of 0 dB, this sensitivity does not play a significant role in spatial localization below approximately 500 Hz.

Nevertheless, localization by means of binaural interaction has two important intrinsic limitations. First, as can be seen by considering the situation in which the space is anechoic and the listener is modeled by a spherical head with ears at the ends of a diameter of the sphere, the interaural parameters (both ITD and ILD) remain constant over any cone around the interaural axis with its apex located at the center of the head, so-called "cones of confusion" (see Figure 2). Thus, for example, under these assumptions, the interaural parameters remain constant (at ITD = 0 and ILD = 0 dB) over all points in the median plane. Second, the interaural parameters convey essentially no information about distance. Only for sources in the near-field[3] do these interaural parameters contain significant distance information.

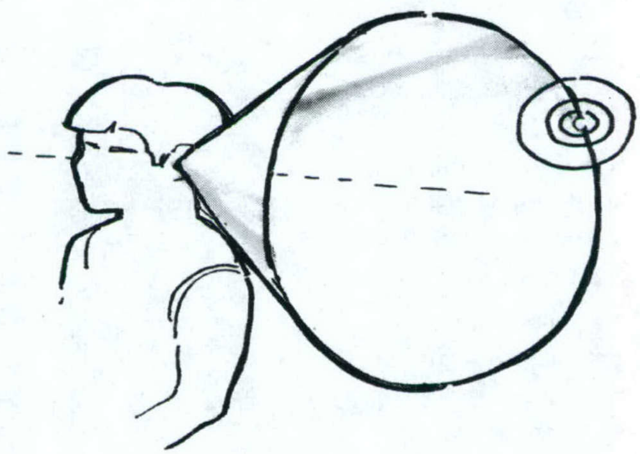


Figure 2.: Cone of confusion. Adapted from [73]

[3] Near-field is the range around the listener where the interaural differences change discernibly when a sound is moved along the radial dimension. The contrary is far-field. A typical near-field envelope radius is about 1 meter. In the far-field the ratio of the distances from the source to the two ears are near unity.

2.1.2.2 Spectral Coloration

To resolve a position on a cone of confusion, it is widely accepted that an additional cue is used. Before reaching the listener's ears, the sound waves are also affected by the interaction with the listener's head, torso, and pinnae, resulting in a directionally dependent spectral coloration of the sound. This systematic "distortion" of a sound's spectral composition acts as a unique fingerprint defining the location of the source. The human brain uses this mapping between spectral coloration and physical location to determine the direction of a sound source.

2.1.2.3 Head-Related Transfer Function

The composite of the ITD, ILD and the spectral coloration characteristics are captured in Head-Related Transfer Functions (HRTF). Even though HRTF's are very rich in acoustic information, perceptual research shows that the auditory system is selective in the acoustic information that it uses in making judgments of the originating direction of a sound source [138]. Due to physical differences between individuals, HRTF's vary greatly in both general shapes and detail [94][96][117][140]. As a result, serious perceptual distortions can occur while listening using HRTF's that were either synthesized or measured on another individual [140][49]. Nevertheless, research shows some individuals experience equal, sometimes improved [25][141], localization accuracy with non-individualized HRTF's – especially when HRTF's of a "good localizer" are used [138].

In general, for each acoustic source in the environment, the signals at the listener's ears can be represented by

$$Y_L(\theta,\phi,d,\omega) = H_L(\theta,\phi,d,\omega)X(\omega)$$

and

$$Y_R(\theta,\phi,d,\omega) = H_R(\theta,\phi,d,\omega)X(\omega), \quad (1)$$

where

(θ,ϕ,d) = spatial coordinates of source relative to the listener's head

θ = azimuth

ϕ = elevation

d = distance

ω = angular frequency

Y_L, Y_R = complex spectra of acoustic signals at the listener's ear drums

H_L, H_R = HRTF

X = complex spectrum of transmitted signal.

Note also that this representation assumes that the source is effectively isotropic [i.e., $X(\omega)$ contains no angular dependence].

2.1.2.4 Localization Process

Given this representation, the process of spatial localization can be described as the process by which the listener determines the spatial coordinates (θ,ϕ,d) from the information contained in the pair of signals $Y_R(\theta,\phi,d,\omega)$ and $Y_L(\theta,\phi,d,\omega)$.

One method for making this determination involves binaural interaction, i.e., comparing the signals at the two ears. For most purposes, this comparison can be represented by forming the ratio

$$\frac{Y_L(\theta,\phi,d,\omega)}{Y_R(\theta,\phi,d,\omega)}=\frac{H_L(\theta,\phi,d,\omega)}{H_R(\theta,\phi,d,\omega)}. \tag{2}$$

Note that forming the ratio eliminates the effect of the transmitted signal $X(\omega)$, that the phase spectrum of this ratio gives the ITD, and that the amplitude spectrum of this ratio gives the ILD. In order to determine the coordinates (θ,ϕ,d) from this ratio, one needs only to know (from previous experience with one's HRTF's) how $H_L(\theta,\phi,d,\omega)/H_R(\theta,\phi,d,\omega)$ depends on (θ,ϕ,d,ω). No knowledge of $X(\omega)$ is required.

A second method for spatially localizing acoustic sources, based not on binaural interaction but on monaural processing, attempts to gain information on $H_L(\theta,\phi,d,\omega)$ and $\mathrm{H}_R(\theta,\phi,d,\omega)$, and thereby on (θ,ϕ,d), by using a priori information on $X(\omega)$ to factor out its influence on $Y_L(\theta,\phi,d,\omega)$ and $Y_R(\theta,\phi,d,\omega)$. Ideally, the system would know $X(\omega)$ well enough to factor its influence out completely, i.e., to form the ratios

$$\begin{aligned} H_L(\theta,\phi,d,\omega) &= Y_L(\theta,\phi,d,\omega)/X(\omega) \\ H_R(\theta,\phi,d,\omega) &= Y_R(\theta,\phi,d,\omega)/X(\omega). \end{aligned} \tag{3}$$

Although the amount of a priori information on $X(\omega)$ available to the listener is seldom adequate to represent the monaural processing in this manner, it is often sufficient to provide reasonably good estimates of $H_L(\theta,\phi,d,\omega)$ and $\mathrm{H}_R(\theta,\phi,d,\omega)$ and thus of some components of (θ,ϕ,d). More specifically, monaural processing is capable of greatly reducing the ambiguities present in the cones of confusion and, in particular, of providing useful estimates of source elevation in the median plane[95]. It should also be noted that listeners who are totally deaf in one ear can still show reasonably good performance in estimating the azimuth of a sound source as well as its elevation.

Generally speaking, the ability of humans to estimate distance is rather poor using either one or two ears. Physical cues relevant to distance estimation include ratio of direct to reverberant energy, overall level, and overall spectral shape. The ratio of direct to reverberant energy and the overall level both tend to decrease with distance, while the friction in air decreases the high-frequency energy with distance, changing the spectral shape. One additional cue to distance that can arise in special cases is how speech is articulated if the talker knows the distance to the receiver.

Finally, and as indicated briefly above in section 2.1.1.2, spatial localization can be substantially altered by the presence of echoes and reverberation. Reflected acoustic energy may have a positive influence on estimation of distance. In certain circumstances, early reflections may enhance localization, but generally, reverberant acoustic energy interferes with spatial localization achieved via either binaural interaction or monaural processing. The degradation it causes in estimation of direction is minimized to some extent by the precedence effect.

2.2 *Altered Hearing*

There are two basic ways in which hearing can be altered: (1) by degradation of biological hearing mechanisms (hearing loss) and/or (2) by introduction of artificial devices or systems that transform acoustic signals prior to biological processing.

2.2.1 Alterations in biological processing (hearing loss)

Although diseases and aging can reduce hearing performance, exposure to noise and loud sounds constitutes a primary cause of both permanent and temporary hearing loss, with perception of high frequencies being particularly vulnerable. The degree and duration of noise-induced hearing loss depends jointly on the level, duration, and spectrum of the exposing sound. For example, continual exposure to impulse noise at 115 dB peak SPL for six hours can produce as much as 60 dB threshold shift for some frequencies, decaying away over the course of days [32]. A dramatic example of profound temporary hearing loss (resulting in permanent hearing loss) caused by a single unprotected exposure to shoulder-borne weapon fire has been recently documented by Vause and Blank [135].

Hearing loss degrades speech reception. The extent and nature of this degradation depends on the degree of hearing loss. People with mild-to-moderate degrees of loss, with pure-tone averages up to about 70 dB, experience difficulty primarily due to inaudibility of the speech signal in at least part of the spectrum [151]. People with losses greater than about 70 dB exhibit some additional speech-reception deficits related to impaired frequency and time resolution, deficits that cannot be compensated with amplification [100].

Hearing loss can also affect sound localization ability. Even when signals are completely audible, some loss in interaural discrimination ability [36] and sound field localization ability [61] is frequently seen with hearing-impaired listeners. However, some listeners with severe loss show no decline in binaural abilities beyond those attributable to audibility [52].

In addition to degradations in detection and localization abilities, hearing loss is sometimes accompanied by tinnitus ("ringing in the ears"). On occasion, this affliction is so severe that total deafness is chosen as a "cure" [71].

Noise-induced hearing loss is, of course, an important problem in the military. In addition to reducing the effectiveness of the warfighter, hearing loss can lead to later costs, both to the government in disability compensation and to the serviceperson in quality of life. Over the past 20 years, hearing conservation programs have made impressive progress reducing the incidence of service-related hearing impairment [105].

2.2.2 Alterations prior to biological processing (head-worn devices)

There are four main categories of devices that alter the acoustic signals reaching the ears. The first category includes devices that provide non-auditory protection (e.g., helmets, goggles, protective bands, etc.). The second includes devices that attenuate the incoming acoustic energy to help protect the listener's ears; the primary types of protective devices are earplugs, earmuffs, and active noise reduction (ANR) muffs. The third category includes head-worn devices that are designed to enhance the listener's hearing; these are primarily hearing aids. The fourth category includes devices that have been developed to provide a combination of hearing protection and enhancement, often called "hear-through" systems. Devices of this type are designed to solve problems similar to those addressed in this project.

2.2.2.1 Non-Auditory Protective Headgear

Few studies have examined localization performance of listeners while using non-auditory protective headgear. Vause and Grantham's study [136] of sound localization included a condition in which subjects wore a Kevlar helmet, which is currently used in the Army. This helmet extends over the ears but does not occlude them, and so does not provide any hearing protection. Subjects localized sounds roughly equally well while using the helmet as when bare-headed, both in the frontal and lateral directions [136].

2.2.2.2 Hearing Protectors

Hearing protectors attenuate the sound reaching the ears to varying degrees depending on the type of device and the care of fitting. Thus, their primary psychoacoustic effect is an increase in absolute threshold. Hearing protectors have little effect on speech intelligibility if the speech signal is strong enough so that it is fully audible after suffering the attenuation of the protector. It is important to note that the audibility limitation becomes important when the user has a significant hearing loss [5].

The results from studies of sound localization with hearing protectors have been consistent with expectations from the known disruption of the physical cues [2][15]. For example, Vause and Grantham [136] showed a large increase in front-back confusions in the horizontal plane with plugs and a Kevlar helmet used together (relative to no device), while errors in the frontal direction were only slightly increased. The authors attributed the increase in front/back errors to the loss of high-frequency spectral cues while using the devices.

When device attenuation becomes very large, even left/right localization can become disrupted [23], an effect that is attributable to mixing of air- and bone-conducted sounds in the cochleae[4]. If the level of the sound conducted by the air path is comparable to that conducted by the bone path, which because of the high speed of sound in bone is effectively the same signal at the two cochleae, then interaural cues will be disrupted. As would be expected, this loss of interaural isolation produces binaural effects like those seen with listeners with conductive hearing losses of about 40 dB or more [148].

2.2.2.3 Hearing Aids

Hearing aids provide amplification to compensate for loss of hearing. Their most important psychoacoustic effects are improved signal detection and speech reception. A frequent negative effect is over-amplification of some sounds, leading to loudness discomfort. To combat this negative effect, and also to minimize the need to adjust the volume control, aids are often provided with some form of automatic gain control. In the most common configuration, independent aids with similar amplification characteristics are worn near, or in, each ear. In response to widespread complaints about hearing-aid amplification of background noise, much recent work has gone into development of microphone arrays that selectively amplify signals from a target direction relative to other directions [57].

Several studies have examined the localization performance of hearing aid users [91][26], using either one or two ear-level aids. Generally, users of binaural hearing aids can localize as well in the left-right dimension with binaural hearing aids as without (with signal level increased to minimize audibility limitations). Some users of monaural aids can localize well in the left/right dimension despite the large asymmetry in levels delivered to the two ears [150]. Sound localization in the median plane is better when the placement of the aid's microphone (e.g., an in-the-ear aid versus a behind-the-ear aid) preserves natural cues [104]. Of course, many hearing-impaired listeners cannot localize well with or without use of hearing aids [61]. Generally, however, the primary concerns of hearing aid research and clinical practice have been on finding the best amplification and compression characteristics for maximizing speech intelligibility and minimizing loudness discomfort; relatively little attention has been paid to localization beyond placement of binaural microphones near each ear to preserve interaural cues for left-right acuity.

Another effect of using a hearing aid is that noise (from the aid's microphone or circuit) can be audible to the user [5]. While this is typically not a major problem with hearing aids because ambient noise usually dominates internal aid noise, it is a potential issue with the proposed Transparent Hearing System when used in very quiet environments.

2.2.2.4 Hear-Through Systems

Hear-through audio devices – also called level-dependent hearing protectors – display the ambient acoustic environment to a listener while also providing protection against strong sounds. They are produced for hunting, tactical surveillance, and military applications. Hear-through devices are the most similar of any head-worn audio system to the proposed Transparent Hearing System. Some types of hear-through devices have the form of protective muffs, while others more resemble earplugs or hearing aids. Earplug types can be further categorized into electronic and passive. Passive level-dependent plugs

[4] The bone-conduction path results from auditory stimulation via conduction of vibratory energy through the torso and skull to the inner ear. Because the bone-conduction path is in parallel with the air-conduction path, sound can be heard via the former path even when the normal air path is completely blocked.

exploit the nonlinear attenuation characteristic of a small orifice [116], and so are also called "perforated plugs". Most electronic versions have a manual volume control to adjust gain for low-level signals and an automatic volume control to reduce gain rapidly for high-level signals.

The main difference between hear-through and transparent hearing systems is the extent to which they are psycho-acoustically transparent, especially with respect to localization based on monaural spectral cues.

There has been little direct research on psychoacoustic effects of hear-through systems, beyond threshold-shift-based measures of attenuation. Measurements of speech reception on some devices [3][4][101] have shown little deterioration in the low-level range; another study [87] found little decrement, relative to open ears, in the ability to identify animal sounds using two types of hear-through devices.

2.3 *Adaptation*

When a listener's hearing is altered by any means, the listener attempts to adapt to the alterations in order to make optimum use of the auditory signals they hear. The extent to which and rate at which a listener can adapt to unnatural auditory signals are important considerations when evaluating auditory systems. For the current project, the need for adaptation will be minimized to the extent that transparency is actually achieved with the Transparent Hearing System. Consideration of adaptation is nevertheless important for two reasons. First, true transparency is not the goal of the Transparent Hearing System. Second, future extensions of the Transparent Hearing System may include processing designed to achieve supernormal listening.

As discussed above, the ultimate goal of a Transparent Hearing System is not to achieve true transparency, because normal human hearing suffers from a wide variety of limitations. Rather, the goal is to achieve the best hearing possible, subject to the constraint that the new hearing provided by the system can be easily learned. The optimal compromise between good hearing performance and short training time has yet to be determined. Knowledge of the human's ability to adapt to alterations of environmental acoustic wave representation clearly constitutes important background information for work in this area.

Although adaptation to altered hearing is clearly a topic of great importance, research in this area has not yet led to adequate understanding or predictive models. Generally speaking, the issue of adaptation or of learning new auditory displays can arise in two contexts: (1) when non-acoustic information is displayed acoustically (e.g., when chemical concentrations or stock market prices are "sonified"), or (2) when the normal auditory representation of acoustic events is altered. Further, within the second context, interest can focus on changes in spatial localization or in changes in other functions of hearing (e.g., speech intelligibility).

2.3.1 Adaptation to Altered or Augmented Hearing

A variety of studies have been conducted specifically to gain better understanding of adaptive mechanisms and limits on adaptation in spatial hearing. Auditory adaptation studies have been conducted to measure a listener's ability to adapt to the use of hearing aids [55], to attenuation of one ear [50] [12], to the use of another individual's pinnae [63], to simulation of a rotation of the ears about the center of the head [62], to simulation of a change in the correspondence between azimuth and spatial cues [121][122][123][124], and to a simulation of increased head size [69]. This area of auditory psychophysics is very complex and is currently receiving considerable attention. No complete summary of human adaptation capabilities or of optimal training procedures to achieve maximum adaptation can yet be constructed. However, some general principles that govern plasticity of the spatial auditory system are beginning to emerge.

2.3.2 Adaptation to Foreign HRTF's

With sufficient experience, a listener can learn to make accurate localization judgments when the correspondence between physical source location and spatial cues is altered. The degree of adaptation that is achieved depends both on the kind of spatial rearrangement and the amount and type of training. Overall, results from a number of studies suggest that feedback and/or interaction with the environment is critical for adaptation [54][63][119][138][140][146]. Further, the amount of training directly affects the severity of the transformation to which a listener can adapt [63].

For instance, there are many common but relatively subtle changes in HRTF's that cause only minor effects on spatial behavior, such as when a listener puts on a hat, changes their head posture relative to body posture, or moves to a different acoustic environment [119]. These results suggest that a listener may constantly "recalibrate" their spatial auditory percepts to overcome minor acoustic distortions and maintain accurate spatial perception as the listener and the environment changes.

With relatively short training periods (on the order of ten minutes of exposure), a listener can learn to overcome some bias in spatial response, provided that the acoustic features that encode source location are grossly similar to those that occur naturally. However, even when short-term training is sufficient to overcome gross localization bias, a mismatch between normal and altered spatial cues can lead to degraded spatial resolution and increased response uncertainty. In general, short-term training with altered cues results in a perceptual "after-effect" whereby listeners exhibit localization bias when presented with normal cues following training with altered cues. In addition, there is evidence that some short-term training effects persist over days, so that users are not "starting from scratch" each time they are presented with altered cues [120][121][122][123][124][125][146].

For more extreme alterations in which the acoustic features encoding spatial location are radically different from normal, short-term training is insufficient, and localization behavior can break down nearly completely [63]. However, with sufficient exposure (e.g. continuous over a period of weeks), even radical alterations of spatial auditory cues can be learned such that response bias is minimal and resolution is equal to or better than normal [63][140]. In addition, with long-term training, both the new and old correspondences between acoustic features and spatial locations can co-exist [63]. In other words, listeners can evidence dual adaptation states, make accurate localization judgments using both normal and altered cues, and switch essentially instantaneously between the two types of cues, as one learns to do with eyeglasses.

Taken together, these studies suggest that short- and long-term training cause qualitatively different perceptual changes. Specifically, long-term training allows the user to learn how to extract and encode new spatial acoustic cues even when these cues are dramatically and qualitatively different from normal cues, essentially learning a new map of spatial cues that does not disrupt the "normal" map. In contrast, short-term experience only can change how the listener responds to a particular set of spatial cues, a change that can cause disruptions in responding to normal spatial cues until the system once again readapts.

THIS PAGE IS INTENTIONALLY BLANK.

3 Current Work

The current work includes a survey of relevant head-borne hear-through auditory systems, a selection of approaches to a transparent hearing solution, implementation of the approaches, and evaluation.

3.1 Survey of Head-Borne Hear-Through Systems

The survey to date has focused on devices that selectively pass-through safe sound while providing protection from harmful noise. A sampling of these devices was studied in detail as part of this project, with results reported in section 5.2.1.

3.1.1 Active In-Canal Hearing Protectors

Active in-canal hearing protectors attenuate sound by blocking the ear canal, while selectively passing-through safe sound filtered by powered means. Because the pinnae are uncovered and the interaural dimension is unaltered, spatial cues are minimally disturbed. The pass-thru frequencies tend to be tightly tuned to speech.

Table 1. Commercial hear-through systems, hearing-aid-in-ear-style.

Manufacturer: Model	Features
Electronic Shooters Protection: ESP-Digital http://www.espamerica.com/products.html	Binaural Mics, AGC, 200hrs, $2000/pr
Micro-Tech: Refuge Hyperacoustic http://www.hearing-aid.com/refuge.htm	Binaural Mics, AGC
Starkey: SoundScope Magnum Ear Digital http://www.earinc.com/p1-electronic-hunting.php	Binaural mics, AGC, 300hrs. battery
Walkers: Digital Game Ear http://www.walkersgameear.com	Binaural mics, AGC, $490/ear
Air Force Communications Earplug (CEP) hear-thru enhancement of Army version	Custom molded, concha and canal plug, ANR, bone-conduction voice mic

3.1.2 Passive In-Canal Hearing Protectors

Passive in-canal hearing protectors filter loud noises while passing-through normal levels by passive means. Because the pinnae are uncovered and the interaural dimension is unaltered, spatial cues are minimally disturbed. Some of these devices place a significant mass in the concha cavity, disturbing these highest frequency cues. The pass-thru frequencies tend to be tightly tuned to speech.

Table 2. Commercial hear-through systems, passive in-canal style.

Manufacturer: Model	Features
Aearo Company: Combat Arms Earplug http://botachtactical.com/aearcomarear.html	Flanged, Dual-use, see Figure 4
Aearo Company: Earlog http://www.aearo.com/html/industrial/earlog3.htm	No battery
Etymotic Research: ER-20 http://www.etymotic.com/	No battery
Jrenum: LD http://www.jrenum.ch	No battery
North Safety Products: Sonic II, Sonic Ear Valves http://www.northsafety.com	No battery
Silencio: Super Sound Baffler FUN-85 http://www.silencio.com/htfiles/earplugs.html	Flanged, see Figure 3
Westone: Style No. 39 http://www.westone.com/earmold_styles.html	Custom molded, concha and canal plug

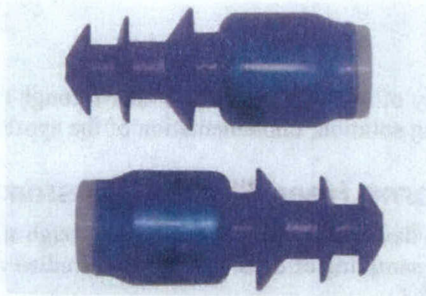


Figure 3: Silencio Super-Sound Bafflers FUN-85 are a good example of a flanged, passively activated in-ear hearing protector. Normal sound pressure levels are passed through its orifice, while a diaphragm is forced closed by high-intensity sound pressures.

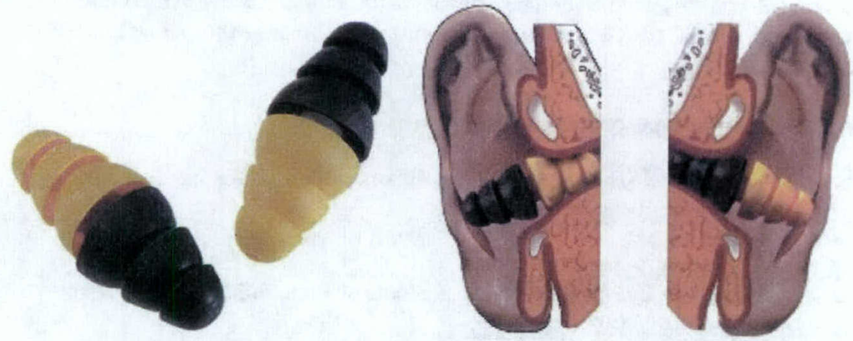


Figure 4: The Combat Arms Earplug is a dual-use device, where one orientation (brown in the ear canal) is a total plug and other (yellow in the ear canal) is a passively activated hear-through protector.

3.1.3 Hunting/Shooting Muffs

There exist currently several commercially available hear-through systems whose goal is to protect human hearing against loud sounds while offering a hearing enhancement system for soft ambient sounds. Most of these systems were designed for individuals who are exposed to high SPL signals and need hearing protection, but who also heavily rely on their hearing for environmental information for situational awareness (e.g. hunters, industrial workers, soldiers).

Separate technologies are used for these dual-purpose systems: one for loud noise suppression, another for hearing enhancement. The hearing protection part of the mechanism usually includes a passive system and an active system, for loud impulse noises. The passive system is composed of sound-attenuating earmuffs that isolate the listener from ambient sound by providing a seal around the ear. At the same time, active electronics detect sudden loud noise and have a limiting system that attenuates the sound to a safe level. The reaction time to sudden onsets is critical as sharp, loud sounds are most harmful to human hearing. The best systems will have a very short reaction time (RidgeLine's ProEars: less than 2 msec).

The hearing enhancement portion of the technology functions on the basis of using a receiver to pickup environmental sounds, amplify them to a comfortable level, and transmit them to the listener. Some products offer a stereo pickup system (ProEars, ComTac, Wolf Ears), while others offer only monaural sound (Detect Ear, Bionic).

Table 3 lists some of the available commercial muff-style hear-through devices along with brief descriptions of features.

Table 3. Commercial hear-through systems, muff-style

Manufacturer: Model	Features
Bilsom: 707 Impact II http://www.bacou-dalloz.com/eu/	AGC, Binaural Mics., 700hrs, Gain Control, water-resistant, $150
Deben: SLIM ELECTRONIC COMMS http://www.deben.com/docs/commshearingURL.gif	Peltor SoundTrap similar
Dillon HP-1 http://www.eguns.com/Dillon_Precision/EyesEars/eyesears.html	Peltor SoundTrap copy, $128
D.P.I. Personal Protective Equip.: Twin Active http://www.dpisekur.com/H.Attive.htm	AGC, Binaural Mics, 50hrs, 16db Gain, rechargeable, balance, Silencio Frontline copy
Gentex: WolfEars http://www.derry.gentexcorp.com/products.htm	Manual level control, AGC, 6db Gain Switch
Howard-Leight: Pro-Ears Leightning http://www.howardleight.com/	AGC, Binaural Mics, Gain Control, $195
Howard-Leight: Pro-Ears Thunder http://www.howardleight.com/	AGC, Binaural Mics, Gain Control
Peltor: ComTac http://www.peltor.com	Binaural mics, manual level control, AGC, Military Grade, 250hrs, $199
Peltor: Sound Trap http://www.peltor.com	Binaural mics, manual level control, AGC, 200hrs
Peltor: Surround http://www.peltor.com	Binaural mics, manual level control, AGC, 100hrs
Peltor: Tactical 7-S http://www.peltor.com	Binaural mics, manual level control, AGC, 100hrs, $149
Pilot Communications: Enhancer (PA 21-10) http://www.pilot-avionics.com/html/hearingprotectorset.htm	AGC, Binaural Mics, 50hrs., 16dB Gain, rechargeable, balance control, speech tuned
Radians Pro-AMP Electronic Muff http://www.botachtactical.com/radproampele.html	Binaural mics Peltor SoundTrap
Silencio: Frontline Electronic HLE-03 http://www.silencio.com	Binaural mics, manual level control, bass/treble, AGC, 50hrs
Silencio: Nighthawk ELP-97 http://www.silencio.com	Binaural mics, manual level control, Balance, AGC, DSP, $180
Silencio: Local HLE-07 http://www.silencio.com	Wireless FRS Comms, Binaural mics, manual level control, AGC, 50hrs
Silencio: Rangesafe Electronic RSX-87 http://www.silencio.com	Monaural mics, manual level control, peakclip, 500hrs, $110
Silencio: Super Ear SSE-01 http://www.silencio.com	Zoom mic, manual level control, no protection, 500hrs
Silver Creek: Bionic Ear http://www.detectear.com	Manual level control, Parabolic Mic, Mono
Silver Creek: Detect Ear http://www.detectear.com	Manual level control, AGC, Parabolic Mic, Mono
Sordin: Supreme III http://www.sordin.com/en/supreme.shtml	Binaural mics, manual level control, AGC, Military Grade, 600hrs
Remington: R2000 Electronic Thin Muff http://www.remington.com	AGC, independent level control $120
RidgeLine: ProEars Dimension http://www.pro-ears.com/	Binaural mics, Independent level control, AGC, 200hrs, $257
Walkers: Power Muffs http://www.walkersgameear.com/quad.asp	Adjustable Attenuation Frequencies, AGC, Binaural Mics, Ind. VC, $259
Walkers: Power Muffs - Quad http://www.walkersgameear.com/quad.asp	Adjustable Attenuation Frequencies, AGC, Quadrophonic Mics, independent level control, $250

Figure 5: RidgeLine's ProEars are a stereo Transparent Hearing System with two microphones mounted flat against the earmuff. The left and right channels have independent manual and automatic volume controls. Loud sounds are attenuated to 70dB with an attenuation attack time of less than 2 msec, while all sounds below 70dB may be amplified up to 70dB.

Figure 6: Gentex's Wolf Ears are a stereo Transparent Hearing System specifically designed for hunters. This system has four main settings: passve hearing protective device only (with a protection of 26 dB), a hear-thru hearing protection device (limiting all sounds to 84 dB SPL), transmission of all sounds at a constant level (automatic gain), and 6 dB amplification boost of all sounds (limited to 90 dB). The left and right speaker channels can be adjusted independently.

Figure 7: Pictured above are four promising COTS active-hear-thru hearing protectors that were not tested. From left to right: Silencio NightHawk, Walker's GameEar Quad, Pilot Communications Enhancer, and Silencio Frontline. Many models are simply private-label copies of other brands, as shown here between the Enhancer and Frontline.

3.2 Approaches

The project scope includes the exploration of a range of approaches to developing transparent hearing solutions. The work is specifically focused on the transparency aspect of the system, i.e., presenting acoustic signals at the ears of the user with occluded/protected hearing such that spatial perceptual accuracy is so well preserved that the user feels confident to perform tasks without removing the protection and does perform tasks requiring spatial awareness equally well. Other necessary components of the complete system, including specific hearing protection, gain control, communications, and supernormal listening, are of secondary interest in the present project. These components were obtained or implemented only as needed to study the transparency approaches.

The primary pathways are shown from one acoustic source to the eardrums of the listener either without any device in Figure 8 (natural hearing) or with a Transparent Hearing System in Figure 9. With natural hearing, the signals at the ears, Y_R and Y_L , result from the source signal being filtered by the pair of head-related transfer functions, H_R and H_L. In the Transparent Hearing System, the source signal is first filtered by the M source-to-microphone transfer functions, P_m. The M microphone signals are then linearly combined by a set of fixed filters, F_{mR} and F_{mL} (not shown explicitly in the Array Processing box) to form the left and right signals Z_L and Z_R delivered to the ears.

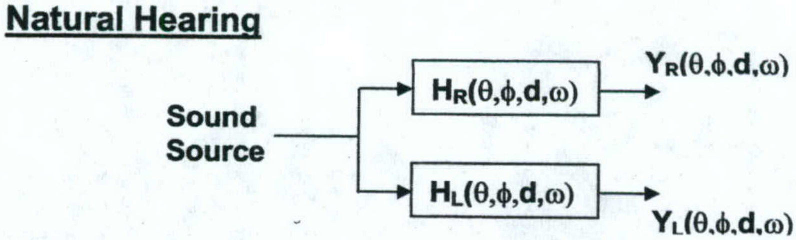


Figure 8: Natural hearing schematic block diagram of the transformation from a signal source to signal spectra at the two ears.

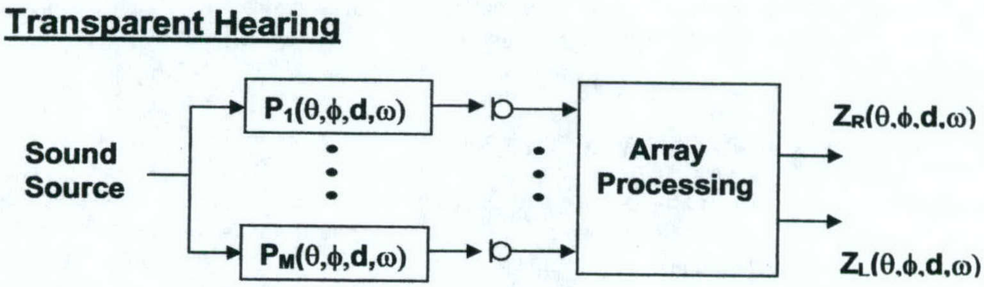


Figure 9: Transparent Hearing schematic block diagram of the transformation from a signal source to signal spectra at the two ears. Note that the Transparent Hearing diagram shows only the components related to achieving transparency. Not shown is an automatic gain control, which would be applied to the Z_R and Z_L signals. Also not shown are the direct acoustic pathways, from the sound source through the hearing protectors and through bone conduction, to the ears. Signals from those paths would mix with Z_R and Z_L.

On the assumptions that microphone transduction and array processing are accomplished linearly and without noise, the transparency goal reduces to the goal of finding the combination of acoustic diffraction functions, *P*, and filters, *F*, that best match the spatial and spectral dependencies of Z_L and Z_R to those of Y_L and Y_R.

Figure 8 and Figure 9 emphasize the two ways in which the spatial and spectral dependencies of the output signals of the Transparent Hearing System can be controlled: 1) by acoustic propagation and diffraction, and 2) by processing multiple microphone signals. Consider, at one extreme, the case of two microphones located on either side of the helmet, with pinna-like structures providing natural diffraction. In this case, the microphone signals themselves will possess the desired dependencies and no subsequent processing would be needed. At the other extreme would be a spatially-distributed array of omni-directional microphones with no diffracting obstacles nearby. In that case, any one microphone signal would have no spatial or spectral dependence, and the desired dependencies would have to be created by filtering and combining the microphone signals.

Metrics are needed to assess the quality of transparent hearing systems. The goal of mathematical equality, $\{Y_L, Y_R\} = \{Z_L, Z_R\}$, would be the basis for the ultimate metric. However, strict equality will be very difficult to approximate, and, given the tolerances in psychoacoustic resolution summarized above, may not be needed. However, it represents one metric for assessing the quality of a prototype solution. Other, less strict, psycho-acoustically-based metrics are also described below and used to guide initial design work. The ultimate test, of course, is functional – does a listener perform auditory tasks as well while using a Transparent Hearing System as they do with no device? Preliminary assessments of prototype systems are part of the work.

3.2.1 Solution Space

The approaches described here for achieving transparency differ along many dimensions, including cost, aesthetic acceptability, and customization capability for individual users. However the primary dimensions reflect the two basic ways of achieving the desired spatial and spectral responses. These two dimensions, labeled "geometric complexity" and "electronic/computational complexity", define a solution space, depicted in Figure 10. "Geometric complexity" means the extent to which the helmet/muff must be modified. "Electronic/computational complexity" includes the simple number of microphones as well as the increased circuitry and processing required.

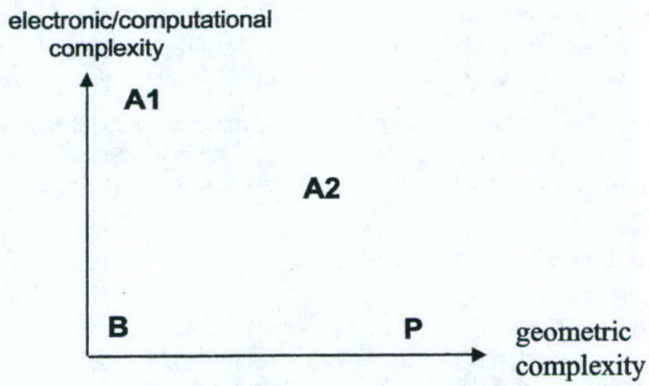


Figure 10: Illustration of the solution space being sampled in this project.

For example, a solution at the point labeled **B** would represent a system that is near zero on both dimensions (e.g., binaural microphones with no added structures). The point **P** would represent something akin to human-like pinna, a solution with no increased processing demands but that requires a special geometric structure. Solution **A1** might be an array of microphones distributed around the helmet that requires no structural modifications. **A2** would be an array of microphones designed to work in conjunction with some added structural elements.

The approaches selected below are an attempt to sample this solution space and to implement in hardware the most promising candidate systems. As an analytical tool in the design stage, an exploration of computational acoustic model methods was conducted in parallel. Computational modeling can potentially enable a quicker and easier sampling of the space prior to implementation than can be achieved with physical models.

3.2.2 Approach Selection

The descriptions in the subsections below present the project teams' selected approaches along with assumptions and prejudices at the onset of this project. Some methods are revealed in this section as part of the approach description.

3.2.2.1 Simple Binaural

The binaural microphone approach represents an elementary receiver system which aims at capturing the fundamental cues used by the human auditory system to localize sounds. The physical setup includes two microphones, placed at either side of the head, with no additional structural elements. The microphone pair is secured in several pairs of locations symmetrically displaced from the median plane. The presence of the head between the microphone pair acts as a natural baffle, leading to inherent ILD and ITD cues. However, due to the unusual shape of the external casing of the helmet and/or muffs, unnatural interaural and spectral properties may result.

There is a tradeoff between the placement of the microphones and the two fundamental cues the binaural microphone approach aims at capturing. Placing the microphones on the ear cups, at the height of the ears, results in a head shadow effect at the contralateral ear, leading to correct ILD cues. Nevertheless, the extended placement of the microphones from the center of the head, in relation to the ears, will lead to exaggerated ITD cues. Conversely, positioning microphones on the helmet, above the ears, may result in accurate time difference cues but may significantly reduce the effect of the head shadow at high frequencies.

The advantage of the binaural microphone system is twofold. First, the simple nature of this approach makes it an easy system to implement and maintain. Second, with few hardware components required to build this system, the total cost will be low. The main disadvantage includes the lack of spectral acoustic information acquired and transmitted to the listener, resulting in a possible loss of localization accuracy. Common to all two-microphone solutions is the limited extension to super-normal listening.

This approach is very similar to that taken by most commercial hear-through systems. However, in this exploration, many important variables are controlled. The primary variables for exploration are 1) microphone location on the headgear, 2) microphone mounting influencing directivity patterns, and 3) microphone isolation.

3.2.2.2 Binaural with Human-like Pinnae

Binaural microphones with human-like pinnae are an extension of the above-described Binaural Microphone approach. The main shortcoming of the Binaural Microphone system is the lack of directionally dependent monaural spectral coloration. In human hearing, these characteristics are created primarily by the cavities in the pinna. The current approach places a pair of microphones mounted in artificial pinnae and positions these on either side of the head, integrated into the protective hearing muff.

Artificial pinnae mounted on a dummy head have been used for many years as a tool for acoustic research as well as by audio engineers for the production of binaural recordings [6][18][46][53][76]. The pinnae are designed and modeled based on the characteristics of human ears and, therefore, accurately simulate the spectral filtration characteristics of real human pinnae. There exist many commercially available artificial pinnae models designed to imitate the ears of humans of different size, age and sex: Knowles Electronic Mannequin for Acoustic Research (KEMAR), Bruel & Kjaer Head And Torso Simulator (HATS), and Neumann KU-100.

The current approach mimics the manner in which human ears receive sound. Its main advantage is that the listener is presented with accurate, natural and complete spatial cues - including ILD, ITD and directionally dependent spectral coloration. Its disadvantages are aesthetic control and position flexibility. Again, common to all two-microphone solutions is the limited extension to super-normal listening.

3.2.2.3 Binaural with Human-like Concha

An approach employing binaural microphones with human-like concha resonance structures is a specialization of the binaural microphone approach. The simpler and less protruding concha structure can be more easily hidden and thus aesthetically addressed than full pinnae structures described in the next sections. This approach will thus be called the "hidden concha" approach in this report to emphasize this advantage. The underlying assumption of a hidden concha system is that source reflections from the ear's concha provide important sound localization cues – especially for indicating elevation. These cues can augment the interaural time and level difference cues provided by binaural microphones on either side of the head that provide for azimuthal localization.

The hidden concha system uses a physical reflecting surface, a 'model concha', to duplicate some of the pinna effects. Figure 11 below illustrates the use of a model of a concha cavity in creating a physical, passive device for replicating spatial cues in a Transparent Hearing System. The key component, the model concha unit, is a physical representation of the concha that reproduces the concha surface and the ear canal accurately. As sounds propagate to the 'ear canal' of the model concha unit, they reflect off the surface of the concha and, as a result, they exhibit useful sound localization cues.

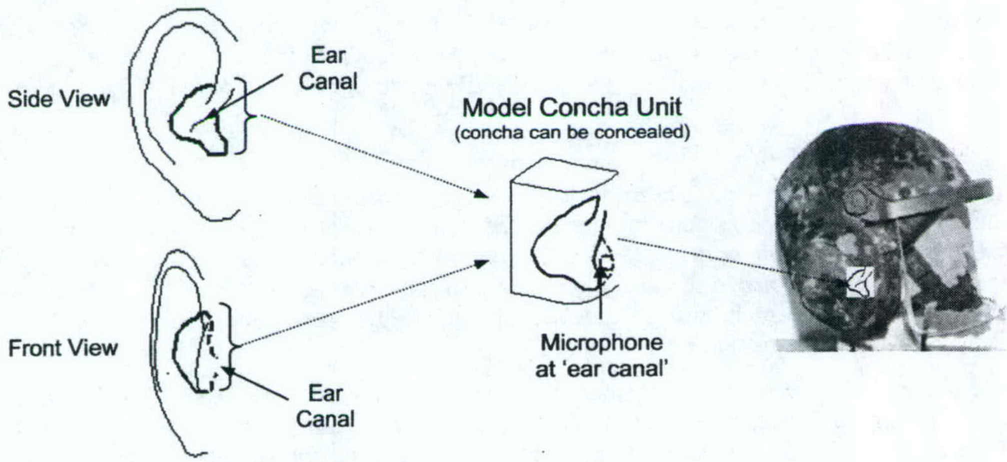


Figure 11: Diagram describing the Hidden Concha System. The smaller concha can provide some of the pinna cues for sound localization while remaining small enough for easy concealment behind a screen or mesh.

The hidden concha approach incorporates left and right model concha units into a selected element of the headgear. The location depends upon the specific headgear constraints. For this study, model conchae were incorporated into 1) the protective hearing muffs with its exaggerated interaural distance, and 2) high on the helmet where an anthropometric interaural distance could be maintained. Exact placement of the model concha units is a significant variable, even within an element such as the hearing muff. For instance, the model concha could have been located towards the front of the muff. Microphones are located in the 'ear canals' of the model concha units. The resulting microphone signals exhibit a combination of binaural cues and concha reflection cues. The binaural cues result from model concha units' location on either side of the wearer's head. The concha reflection cues provide the wearer with a more realistic sense of space than having no resonance chamber at the transducer.

An important feature of these model concha units is their size: they are smaller than the pinna as a whole. This smaller size allows the model concha itself to be 'hidden' behind a mesh or screen in the model concha unit. This is an important advantage of hidden concha systems, since the concealment of the concha surface liberates the system design from some aesthetic concerns. This freedom means that the model conchae can be as realistic as possible to provide maximal acoustic transparency. In fact, the model conchae could even be personalized using ear-molds for each soldier, which may further improve performance. Additionally, the model concha units could be interchangeable between various sets of protective muffs: once a soldier has a set of model concha units made, it would fast and easy to personalize any given transparent audio system.

The main disadvantage of hidden concha systems is that the concha is only part of the whole pinna. Without adaptation, concha reflections may be insufficient to augment the binaural microphone system performance to the desired levels of acoustic transparency. This approach contrasts to commercial in-ear systems that fill the concha and depend on the outer pinna cues exclusively. Once again, common to all two-microphone solutions is the limited extension to super-normal listening.

3.2.2.4 Binaural with Mechanically-Modeled Pinnae Cues

This approach explores geometric shapes, integrated into the helmet and protective muff design, which show promise to convey signals to binaural microphones with direction-dependent spectral colorations, thus simulating pinnae cues.

While at first glance the helmet's function appears to be primarily ballistic protection, in fact it serves both as a protection and a sensing platform. As such, the task of outfitting the helmet with acoustic sensors requires a number of considerations across different disciplines. A subset of these issues is presented here.

Acoustic Characteristics

Just as a helmet affects the sounds heard inside the helmet, it also effects how sound is filtered near the outside surface of the helmet, where microphones are likely to be placed. As such, the basic form of the helmet, and all associated gear mounted to it, changes the basic acoustic field that is detected by the microphones. When coupled with microphone placement, these effects can be good: simplify processing and augment hearing, or can be bad: occlude acoustic information that cannot be recovered, or increase the computational signal processing requirements.

It is useful to note that there are two basic levels on which form design and exploration is key. Macro(form), which considers the overall shape of the resultant helmet design, and Micro(form) which considers the acoustic cavities that might be employed to cradle individual microphones and generate spatial cues.

Aesthetic Considerations

Physical forms which anthropomorphically resemble humans, run an aesthetic risk of becoming caricatures of the human features they resemble. The mechanical modifications made to the helmet must be such that the user will be proud to wear it.

Shape Singularities

The helmet will be used in adverse environments where it will be subjected to obstacles that can snag (like twigs or brush) or to liquids (like water or chemical coverings). A well-designed helmet will not snag or collect debris in any number of adverse environments.

Human Factors

Stability, fit and comfort must not be negatively affected by the modifications made for acoustic considerations. An obvious concern is the addition of mass or the relocation of the helmet's center of mass, with particular attention paid to rotational moments of inertia [110]. Another critical concern is heat and perspiration dissipation, as the ears are cooling radiators for an overheated human.

Modularity

The helmet is generally a modular protection and sensor package. As such, any new mechanical designs must be optimized so as not to overly reduce the ability of the helmet to be outfitted with different types of sensor and protection devices. For example, perhaps a microphone will need to be covered to allow the helmet to accept a new sensor or processing module.

As an example of this exploration, Figure 12 shows the Natick "Scorpion R2" helmet design. This helmet design has received some accolades for "looking cool" as well as providing adequate function. The design features many direction-dependent crevasses that lend themselves to being leveraged or modified to provide the effects of human pinnae and conchae.

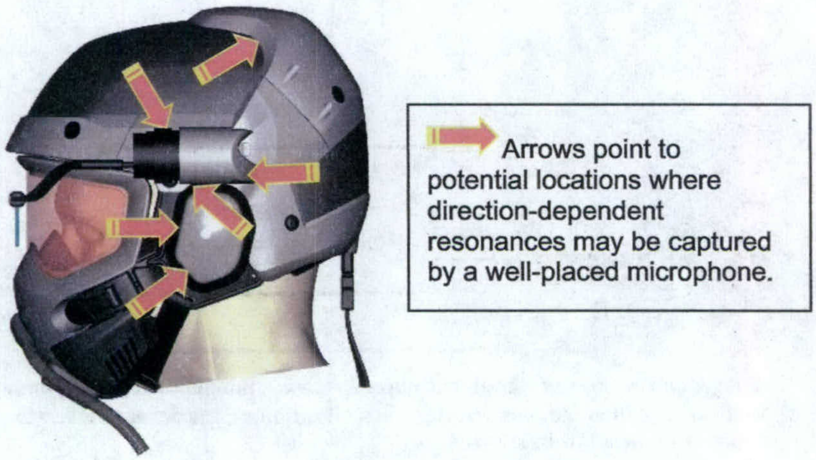


Figure 12: Natick "Scorpion R2" helmet design.

This approach provides the advantage of minimal processing cost, while still delivering some spatial cues with a desirable aesthetic. The potential disadvantage is that the spatial cues might be non-optimal. This approach may not lend itself well to future supernormal listening capabilities.

3.2.2.5 Pinna-Simulating Clustered Array

The simulated-pinnae approach is another generalization of the binaural-microphone approach. Instead of using a physical structure to duplicate the pinnae localization cues, however, this approach attempts to duplicate the pinnae cues using microphone array processing. It operates by replacing the binaural microphone pair with two small clusters of microphones (e.g., 2-4 mics per cluster) located near the left and right ears of the soldier. Localization cues are provided by a combination of microphone-cluster placement on either side of the head (binaural cues) and array processing (pinnae cues). As stated above, the placement of the simulated-pinnae microphone clusters is selected to approximate the desired HRTF binaural cues. The simulated pinna microphone-array processing designs concentrate on reproducing the magnitude response of the monaural spectral pinna cues. The augmentation of binaural cues with magnitude-response pinnae cues should provide enhanced localization cues and should increase the transparency of the system. Since the left and right ear processing are identical, the following discussion presents the simulated pinnae system for a single ear.

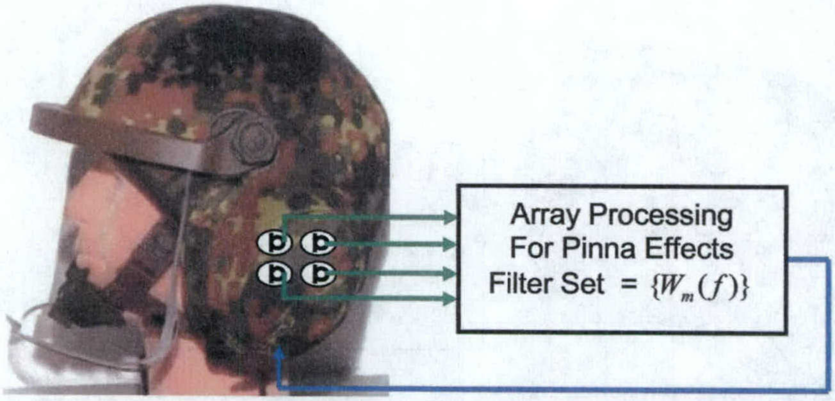


Figure 13: Diagram showing the Simulated Pinna System. Independent microphone arrays located at each ear provide pinna localization cues. Positioning the arrays on either side of the head provides binaural localization cues.

General Optimization Simulated-Pinnae Systems

Figure 13 depicts the basic idea behind using an arbitrary microphone array system to generate the pinna cues for a single ear. As discussed in Appendix A, an M-microphone array system has a directional response that is governed by the combination of microphone placement and microphone-array filter selection. Specifically, the directional response:

$$G(f,\theta,\phi)=\sum_{m=1}^{M}H_m(f,\theta,\phi)W_m(f), \tag{4}$$

where $H_m(f,\theta,\phi)$ is the transfer function from a source at (θ,ϕ) to the m^{th} microphone and $W_m(f)$ is the filter applied to the m^{th} microphone signal. The microphone array can then be used to simulate pinna cues by selecting array-processing filters so that $G(f,\theta,\phi)$ is as close as possible to the desired directional pinna response.

For a given set of microphone locations, the process of selecting the filter set $\{W_m(f)\}$ to yield the desired $G(f,\theta,\phi)$ is conceptually straightforward, since the dependence of $G(f,\theta,\phi)$ upon $\{W_m(f)\}$ is very explicit. Given the measured $\{H_m(f,\theta,\phi)\}$ and a desired pinna directional response $P(f,\theta,\phi)$, the most direct method to choose $\{W_m(f)\}$ is to minimize the squared error between $G(f,\theta,\phi)$ and $P(f,\theta,\phi)$ over the location set of interest. This results in the Least Squared Error (LSE) solution:

$$\{W_m(f)\}_{\text{lse}}=\underset{\{W_m(f)\}}{\operatorname{argmin}}\sum_{(\theta,\phi)}w_{\text{lse}}(f,\theta,\phi)\,|\,G(f,\theta,\phi)-P(f,\theta,\phi)\,|^2, \tag{5}$$

where a location-dependent weighting term $w_{\text{lse}}(f,\theta,\phi)$ has been included so that the array directional response can be made more accurate for higher-importance spectral features such as the elevation-dependent notches evident in most HRTFs. Note that both $G(f,\theta,\phi)$ and $P(f,\theta,\phi)$ are complex valued and that the error in this solution is a complex-distance error and accounts for both magnitude and phase.

The main advantage of the LSE solution $\{W_m(f)\}_{\text{lse}}$ lies in the fact that it has a simple, closed-form solution at each frequency for a discrete set of locations (θ,ϕ). Once $\{W_m(f)\}_{\text{lse}}$ has been determined at each frequency, FIR approximations can be determined and a system can be designed. The main disadvantage with this solution is that its error definition tends to be too general. As stated above, the simulated-pinnae systems should be designed to reproduce the magnitude response of the pinna spectral cues. The LSE error definition tries to match both the magnitude and phase of the desired pinna cues. The inclusion of phase information in this optimization can significantly alter and limit the ability of the LSE approach to match the pinna magnitude response.

For this reason, an alternative simulated-pinnae design method focuses on the preservation of only the pinna-cue magnitude response. In the Least Square Magnitude Error (LSM) solution, the set $\{W_m(f)\}$ is chosen to minimize the squared error between $20\log|G(f,\theta,\phi)|$ and $20\log|P(f,\theta,\phi)|$ over the location set of interest:

$$\{W_m(f)\}_{\text{lsm}} = \underset{\{W_m(f)\}}{\operatorname{argmin}} \sum_{(\theta,\phi)} w_{\text{lsm}}(f,\theta,\phi)(20\log|G(f,\theta,\phi)| - 20\log|P(f,\theta,\phi)|)^2, \tag{6}$$

where $w_{\text{lsm}}(f,\theta,\phi)$, like $w_{\text{lse}}(f,\theta,\phi)$ in the LSE solution above, is a location-dependent weighting term that emphasizes more important desired HRTF features.

Simplified Two-Microphone Delay-and-Sum Simulated Pinnae Systems

The LSE and LSM simulated pinnae optimizations outlined above are intended for generating pinna cues from arbitrary microphone clusters. It is possible, however, to create somewhat simpler systems that still preserve some significant features of the pinna magnitude response. Specifically, consider the elevation-dependent notch evident in the desired HRTF magnitude response of Figure 14. Such a notch can be generated simply and effectively using the two-microphone Delay-and-Sum simulated pinnae architecture shown in Figure 15.

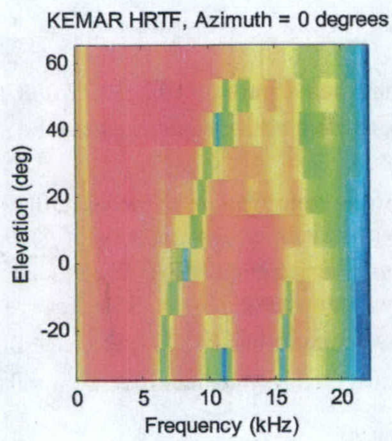


Figure 14: An example from the desired HRTF dataset, showing 10 elevations at 0 degrees azimuth HRTF's in magnitude spectral plot. This data was taken from KEMAR[53].

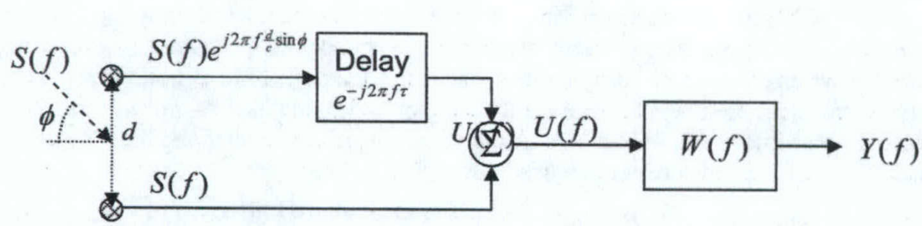


Figure 15: Architecture a two-microphone delay-and-sum simulated pinnae system.

To see how this notch is generated, consider two free-field microphones oriented vertically and spaced d m apart. Given a source $S(f)$ arriving to the lower microphone from an elevation ϕ, then the top microphone input is a time-advanced copy of the lower microphone input:

$$S(f)e^{j2\pi f\frac{d}{c}\sin\phi},$$

where c is the speed of sound. Delaying the upper microphone signal by τ sec and summing with the lower signal leads to the intermediate result:

$$U(f)=S(f)[1+e^{j2\pi f(\frac{d}{c}\sin\phi-\tau)}]. \quad (7)$$

This signal exhibits a null at:

$$f=\frac{-1}{2(\frac{d}{c}\sin\phi-\tau)}. \quad (8)$$

For $d=0.01\text{m}$ and $\tau=70\,\mu\text{sec}$, this null occurs at 5918, 7143, 9008, and 11136 Hz for ϕ equal to -30, 0, 30, and 60 degrees. This null variation with elevation mimics the null variation seen in the desired HRTF shown in Figure 14.

This elevation-dependent null is only one major feature of the desired HRTF, however. The final output of this simplified simulated-pinnae system is formed by passing $U(f)$ through the filter $W(f)$. This filter is designed to account for the remaining desired HRTF features and is chosen in a manner similar to that used in the LSM simulated pinnae optimization above. Specifically, $U(f)$ is regarded as the single-microphone input to a simulated pinnae system and $W(f)$ is the LSM filter from Equation (6) above that minimizes the magnitude error between the system output $Y(f)$ and the desired HRTF.

Collective Simulated-Pinnae Notes

Simulated pinnae systems in this study use the preceding techniques to determine the microphone placement and array-processing filters for each ear. The array configuration and processing filters are determined at design time and will not be updated actively while the system is in use. As stated above, the microphone array in this system is optimized primarily to generate appropriate pinna cues in the outputs for each ear. The positioning of the left-ear and right-ear microphone arrays on either side of the head provide the interaural time-delay and level difference cues that are also important in sound localization.

The simulated-pinnae system has the following advantages. First, it uses no physical reflecting surface to provide the simulated pinna cues. This means that there are essentially no aesthetic concerns over the appearance of the systems, since the microphones in each cluster are easily concealed. Second, the software-based nature of this system allows for great flexibility in the system design. The array-processing filters can be potentially designed to approximate any HRTF's – including generic HRTF's from a database, measured HRTF's for the soldier actually using the system, or enhanced HRTF's designed to improve soldier performance.

3.2.2.6 Sound-Field Microphone

As stated previously, the goal of transparent hearing is to capture the sound-field arriving at the listener and to accurately reproduce and present the sound to the listener's ears (around obstacles such as headgear and hearing protection) in a way that preserves location information and feels natural. By using a sound-field microphone[5], it is possible to capture the three-dimensional sound field and present it using headphones to a listener. If it were possible to create a sound-field microphone on or around the helmet or headphones, that sound-field could be converted to a binaural signal, thus giving the listener a natural display of the sound-field that preserves direction information. In addition, virtual 3-D sources (such as communications signals) can be efficiently encoded into Ambisonic B-format[5] and mixed in with the microphone signal. Therefore, it would not require any additional filtering resources to have a mixture of virtual and actual sources presented to the listener.

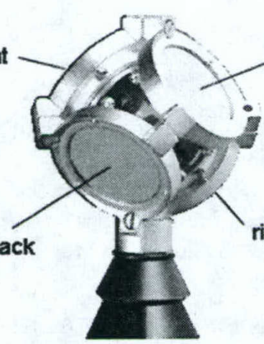


Figure 16: An example sound-field microphone capsule.

Sensing a sound-field around an object

The sound-field microphone consists of four cardioid microphones mounted in a tetrahedron (see Figure 16). Ideally, the microphones would be coincident, but since that is not possible they should be mounted as close as possible. By using small electret microphone capsules, it is possible to make a sound-field microphone that would be between ½" to 1" in diameter. The further apart the microphones are, the less accurate the captured sound field will be. In addition, sound-field microphones are designed to work in the free field. These restraints make it difficult to have an accurate sound-field capture at the listener. Several possibilities were explored:

- Designing a sound-field microphone around the head. If possible, this could capture the sound field arriving at the head. However, it is likely that the necessary distance to the capsules would make the error too large to do this with just 4 microphones. It might be possible with a larger number of microphones, but much of the simplicity of Ambisonics would be lost.
- Placing a sound-field microphone on the top of the head. This would have good performance for sounds that are not close up. Near-field sounds would be distorted, because of the difference in height between the center of the head and the top of the head. In addition, there

[5] A brief overview of sound-field microphones and Ambisonic theory is given in Appendix B.

would be a shadowing of sounds coming from below and a possible reflection of sounds coming from above off the helmet that would make them sound like they were coming from below.

- Placing a sound-field microphone at each ear. This has the advantage of generating a much more accurate sound-field to each ear since the sound field very far from the captured location would not be extrapolated. It has the disadvantage of requiring twice the number of channels, and does not support rotating the sound-field after capturing the sound[6]. It might be possible to reduce the number of microphones needed to three per ear instead of four by orienting the tetrahedron such that one side is flat against the helmet, and making the assumption that the microphone that should be there is totally occluded by the helmet.

After the microphones are built and mounted on the helmet, it is necessary to convert the microphone signals into B-Format. For a tetrahedral sound-field microphone, label the four microphones Lb, Lf, Rb, and Rf, depending on their Left, Right, front and back orientations as shown in Figure 16. If the 4 microphone capsules are exactly coincident, then the conversion is the simple linear combinations [98]:

$$
\begin{aligned}
W &= Lf + Rb + Rf + Lb \\
X &= Lf - Rb + Rf - Lb \\
Y &= Lf - Rb - Rf + Lb \\
Z &= Lf + Rb - Rf - Lb
\end{aligned}
\tag{9}
$$

However, since the microphone capsules cannot overlap and must be offset slightly from each other, the conversion must be corrected for this separation. This can be accomplished by measuring the impulse response of the microphones in different directions, and setting up a system of linear-filtering equations that can be solved with least-squares or other numerical methods. For example, the B-Format signals could be formed using linearly-filtered additive combinations of the four sound-field microphone signals [98]:

$$
\begin{aligned}
W &= h1w \otimes Lf + h2w \otimes Rb + h3w \otimes Rf + h4w \otimes Lb \\
X &= h1x \otimes Lf + h2x \otimes Rb + h3x \otimes Rf + h4x \otimes Lb \\
Y &= h1y \otimes Lf + h2y \otimes Rb + h3y \otimes Rf + h4y \otimes Lb \\
Z &= h1z \otimes Lf + h2z \otimes Rb + h3z \otimes Rf + h4z \otimes Lb
\end{aligned}
\tag{10}
$$

Note: since this approach does not require specific microphone geometry, one benefit is that it could create B-format signals from any arbitrary microphone array and not just from a sound-field microphone. Additionally, this approach accounts for the differences in microphone capsule responses, which makes it less critical to find perfectly matched microphone capsules.

After the A-format from the microphones is converted to B-format, the next step is to decode the B-format signal. There are two approaches towards creating "Binaural B-Format". One is to use conventional Ambisonic decoding for a speaker array and then render that array with "virtual speakers" inside a simulated environment [88][132]. To get good spherical coverage, at least 12 virtual speakers should be employed, possibly arranged in an icosahedron or other regular polygon. Alternatively, it is possible to convert from B-format to binaural by projecting the spherical harmonic basis set onto the HRTF data set [43].

Microphone Array to HRTF Transfer Functions

Ultimately, the goal is to transcode from the microphone inputs to the two channels of a binaural mix. While B-format is a useful intermediate representation that allows for some efficient manipulations of the

---

[6] The rotation is only useful for spatially-rendering the sound-field to remote listeners.

sound-field (rotation in particular), future work should explore the design of a more efficient and accurate transformation that bypasses B-format. The goal of such an alternative transformation is to find a set of filters that optimally converts a set of microphone signals so that they exhibit the desired HRTF responses. In the case of a sound-field microphone, left and right output is formed by applying left and right sets of filters to the sound-field microphone signals and summing the results:

$$L = hl_1 \otimes Lf + hl_2 \otimes Rb + hl_3 \otimes Rf + hl_4 \otimes Lb$$

$$R = hr_1 \otimes Lf + hr_2 \otimes Rb + hr_3 \otimes Rf + hr_4 \otimes Lb \qquad (11)$$

More generally, given an arbitrary set of microphones M_n, the left and right outputs would be:

$$L = \sum_n hl_n \otimes M_n$$

$$R = \sum_n hr_n \otimes M_n \qquad (12)$$

The filters hl_n and hl_n are determined by measuring the microphone input responses from several different directions using a least-squares approach to determine the filters that most effectively convert the microphone signals into the desired measured HRTF responses for those directions. The least-squares optimization could also include regularization, which employs frequency dependent weighting in the optimization [131][74]. This regularization produces a more optimal solution by applying heavier weight to frequency bands that are known to be more accurately measured and known to be most important to the sound localization.

3.2.2.7 General Microphone Array

The general microphone array approach is similar to the simulated pinnae approach described above and the distributed array described later in that it uses an array of microphones to generate the HRTF cues. The important differing factor from the simulated pinnae approach is its use of one single large array to generate both the left and the right output signals of the system. Recall the simulated pinnae system uses two, small, independent arrays to produce the left and right ear outputs separately. In contrast to both the distributed array and the simulated pinnae, this approach does not depend on specific physical microphone placement for cue preservation. In addition, the binaural systems all use the presence of the head to provide the desired binaural time and level difference localization cues, while depending on other independent means (model pinnae, model conchae, clusters of microphones,...) to generate the spectral pinna localization cues. Because general microphone array systems use a single array to generate both output signals, care must be taken to preserve both binaural and spectral HRTF cues. Figure 17 shows the basic structure of the general microphone array architecture. For this system, several microphones are mounted throughout the assembly. All microphones are passed to two separate array-processing systems that generate the left and the right ear signals, respectively.

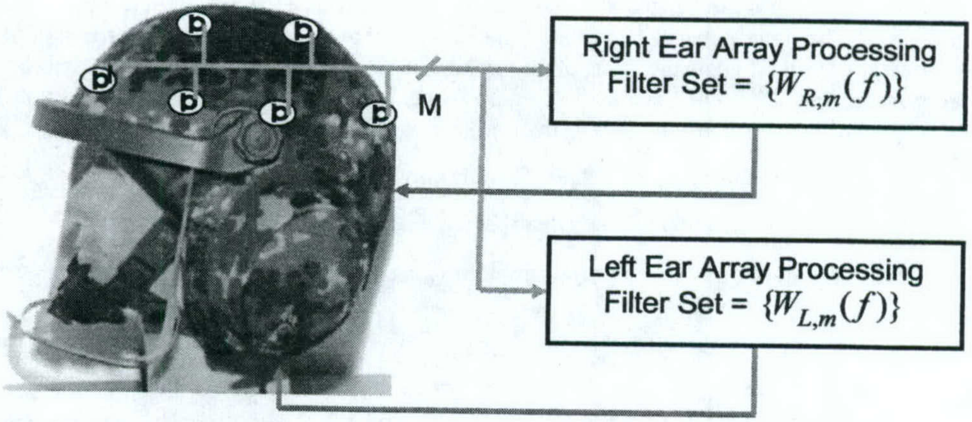


Figure 17: Diagram depicting the General Microphone Array approach. A microphone array distributed over the helmet is used to synthesize approximate HRTF's. The system is designed to generate both pinna and binaural localization cues.

The goals of preserving the binaural and spectral HRTF cues can be difficult to meet simultaneously, thus the general microphone array systems developed in this work separate the two goals. Binaural cues (in particular interaural time differences) are typically most important at low and mid frequencies, while spectral cues (such as notches) are most important at high frequencies. Given this knowledge, the general microphone array systems developed in this work concentrate on binaural cue preservation at frequencies below 4kHz and on spectral cue preservation at frequencies above 4kHz.

Binaural Cue Preservation

Binaural cues are preserved by identifying two reference microphones, one 'left' and one 'right', from the array based upon their proximity to the true ear locations. This selection of microphones ensures that the inter-microphone time and level differences are similar to the natural interaural ones. These left and right reference microphones are then treated as single-microphone simulated pinnae systems. Microphone filters $W_{L,\mathrm{bin}}(f)$ and $W_{R,\mathrm{bin}}(f)$, are generated based on

Equation (6) $G_R(f,\theta,\phi)=\sum_{m=1}^{M} H_m(f,\theta,\phi)W_{R,m,\mathrm{spec}}(f)$, (13) above to equalize the average microphone magnitude responses to the desired left and right HRTF magnitude responses.

Spectral Cue Preservation

Spectral cues are preserved by using the entire array to generate two outputs that minimize the error between the left and right HRTF responses. Given left and right array processing filter banks, $\{W_{L,m,\mathrm{spec}}(f)\}$ and $\{W_{R,m,\mathrm{spec}}(f)\}$, and using the array-processing concepts put forward in Appendix A, it is possible to express the left and right directional responses of the array as functions of the array filters and source location:

$$G_L(f,\theta,\phi)=\sum_{m=1}^{M} H_m(f,\theta,\phi)W_{L,m,\mathrm{spec}}(f),$$

$$G_R(f,\theta,\phi)=\sum_{m=1}^{M} H_m(f,\theta,\phi)W_{R,m,\mathrm{spec}}(f), \qquad (13)$$

where the $H_m(f,\theta,\phi)$ are the measured (and time-invariant) source-to-microphone frequency responses as functions of frequency and location. Given $G_L(f,\theta,\phi)$ and $G_R(f,\theta,\phi)$ and the desired HRTF values $P_L(f,\theta,\phi)$ and $P_R(f,\theta,\phi)$, the most straightforward way to match the spectral cues in the desired HRTF is to select $\{W_{L,m,\text{spec}}(f)\}$ and $\{W_{R,m,\text{spec}}(f)\}$ to minimize the squared error between these values over the location set of interest. This results in the general microphone array interpretation of the simulated-pinna least squared error (LSE) solution presented in Equation (5) above:

$$\{W_{L,m,\text{spec}}(f),W_{R,m,\text{spec}}(f)\}_{\text{LSE}} = \underset{\substack{W_{L,m,\text{spec}}(f)\\ W_{R,m,\text{spec}}(f)}}{\text{argmin}} \sum_{(\theta,\phi)} w_{\text{LSE},L}(f,\theta,\phi)\,|\,G_L(f,\theta,\phi)-P_L(f,\theta,\phi)\,|^2 + w_{\text{LSE},R}(f,\theta,\phi)\,|\,G_R(f,\theta,\phi)-P_R(f,\theta,\phi)\,|^2, \tag{14}$$

where $w_{\text{LSE},L}(f,\theta,\phi)$ and $w_{\text{LSE},R}(f,\theta,\phi)$ are location-dependent weighting terms that serve to enhance important spectral cues in the optimization.

As with the simulated-pinnae LSE solution, the main advantage of the general microphone array LSE solution lies in the fact that it has a simple, closed-form solution at each frequency for a discrete set of locations (θ,ϕ). The main disadvantage with this solution is that its error definition tends to be too general, since it attempts to preserve both the magnitude and phase of the desired HRTF. Spectral cues require preservation of only the magnitude response of the desired HRTF. While it is possible to formulate a general microphone array interpretation of the simulated-pinna least squared magnitude (LSM) approach described in Equation (6) above, practical experience has shown that this optimization does not converge well for systems with more than 4 microphones. Since most general microphone array systems consist of more than 4 microphones, a general microphone array LSM solution is not considered in this research.

Once $W_{L,\text{bin}}(f)$, $W_{R,\text{bin}}(f)$, $\{W_{L,m,\text{spec}}(f)\}$, and $\{W_{R,m,\text{spec}}(f)\}$ have been determined, the final general microphone array filters are generated by applying $W_{L,\text{bin}}(f)$ and $W_{R,\text{bin}}(f)$ to the appropriate lowpass-filtered left and right reference microphones and adding these results to the highpass-filtered $\{W_{L,m,\text{spec}}(f)\}$, and $\{W_{R,m,\text{spec}}(f)\}$ outputs from the entire array. This leads to final left and right ear system filters of:

$$W_{L/R}(f) = \begin{cases} \text{LPF}(f)W_{L/R,\text{bin}}(f) + \text{HPF}(f)W_{L/R,m,\text{spec}}(f), & m = L/R \text{ reference mics,} \\ \text{HPF}(f)W_{L/R,m,\text{spec}}(f), & \text{otherwise,} \end{cases} \tag{15}$$

where $\text{LPF}(f)$ and $\text{HPF}(f)$ are lowpass and highpass filters with cutoffs at 4kHz.

General Microphone Array Notes

As with the simulated-pinna, sound-field microphone, and distributed array systems, the microphone configuration and the array-processing filter sets for the general-array system are determined at design time and are not updated (non-adaptive) while the system is in use.

Assuming a reasonable optimization can be found, this approach has advantages similar to those of the simulated pinna systems: specifically, concealed microphones do not violate any aesthetic constraint that might exist for the system, and software-based processing allows for the customization of the system directional response to approximate any HRTF's (generic, custom-measured, or enhanced). The general microphone array system has additional advantages of microphone placement flexibility, since specific placement is not required, and extensions to super-normal listening capabilities.

3.2.2.8 Distributed Array with 3D Processing

From the body of knowledge surrounding spatial cue synthesis, it is known that from each bearing of sound arrival to a listener, there is a distinctive head-related transfer function which transforms the sound from a free-field wave to respective binaural signals entering the left and right ear canals. This simple relationship can be mapped to the whole sphere around the listener using superposition of the linear system. The theory, however, is true for the limit of infinitely separating each incidence of sound wave direction, but is an approximation for less than the limit. Further, the theory assumes a means of independently and exclusively sensing each incident direction. Still, these collective approximations may be psycho-acoustically better than the approximations of other approaches.

The major obvious drawback/tradeoff of this approach is expense. It requires significantly more microphones than any other alternative approach under consideration in this project. It requires fixed-filter HRTF processing, which, although computationally an order of magnitude cheaper than interactive HRTF processing, is very processor intensive. The microphones are distributed evenly over the entire surface of the helmet in a polyhedral pattern, leaving little, if any, place to attach helmet accessories without disturbing the sensor array's performance. Basic filtering can compensate for invariant disturbances of known accessory configurations. But this is again more expense in processing.

Simple microphones are generally more omni than directional. Each microphone, augmented by well-designed acoustic coupling, can have a principal directionality, but will not provide flat direction exclusivity. Psycho-acoustically, this flaw across an array of microphones responding to the same stimulus can result in a blurring of the perceived direction. The microphone directionality can be sharpened by using array processing techniques (beam-forming) with neighboring microphones. Again, this is yet more expense from additional processing.

In summary, if expense was not a factor, this approach yields excellent transparent performance and provides an optimal platform for supernormal listening. It is a brute-force approach akin to the sensors coating a fly's eye. This approach differs from the general microphone array in that it requires directional coverage (distribution) with the microphone and thus promotes assumptions that circumvent optimization steps for filter design.

4 Methods and Implementation

This section describes the specific systems evaluated in this study, the methods for design, how they were implemented, and the means by which their performance was measured.

4.1 Theoretical and Numerical Modeling

Before physical implementation of the approaches described in the previous section, many were studied theoretically to determine a preliminary evaluation of their respective levels of effectiveness in achieving the goal of maximal acoustic transparency and to guide in the initial selection of parameters for physical prototyping. The original proposal for this study had prospected that numerical modeling tools could be used in the theoretical analysis, yielding usable results for both this exploration and future headgear system designs. However, the numerical modeling did not bear early fruit during the project, and thus became a parallel study of its own.

The objective was to employ computational models of sound propagation to virtual microphone locations on the surface of a three-dimensional geometric model representing a human head and helmet. These models would attempt to include various aspects of the helmet itself: e.g., the protective ear muffs and the scopes, sensors, and devices mounted on the helmet. Additionally, for the binaural-microphone approaches, the models would incorporate pinna feature approximations.

The goal of this modeling was to help identify the pinna structures and the microphone placements, for both binaural-microphone and multi-microphone approaches, that will yield the most acoustically-transparent system. These would lead to faster initial evaluations in the design stage than can be achieved with physical models.

The computational modeling of simplified models of the head and torso (such as two rigid spheres depicting the head and torso respectively) has led to reasonable approximations and has a relatively light computational burden [7]. Subsequently more complex geometries to include more realistic features on the helmet (e.g., stylized pinna structures, and helmet sensors/components) may be simulated using the boundary element method (BEM) to numerically solve the partial differential equation (PDE) governing the sound-pressure field at selected microphone locations. The boundary element class of algorithms [75] is generally regarded as computationally efficient for this sort of acoustic scattering problem, although the level of model complexity that this method can simulate with a reasonable amount of computation remains an open question.

Since the numerical methods were not in themselves employed in analyzing any of the Transparent Hearing System approaches, the theoretical methods employed were not rigorous, but rather speculative. For a full description of the methods employed for numerical modeling, please see the results section.

4.2 Physical Prototyping

Nine independent physical prototypes were constructed, seven of which were completed to a human-wearable form. Leveraging those seven physical prototypes, 56 distinct variations relating to microphone placement and other configuration differences are testable.

4.2.1 The Helmet and Muff platform

For consistency, the CGF/Gallet TC-2001 Sidecut MICH helmet coupled with Sennheiser HD-205 passive-attenuating headphones has been selected for the base platform of all approaches, shown below in Figure 18. The TC-2000 MICH helmet is the current new standard under steady adoption by many warfighting groups. The Sidecut MICH provides the same basic shape of the MICH, but with clearance for exposing large passive-attenuating muffs.

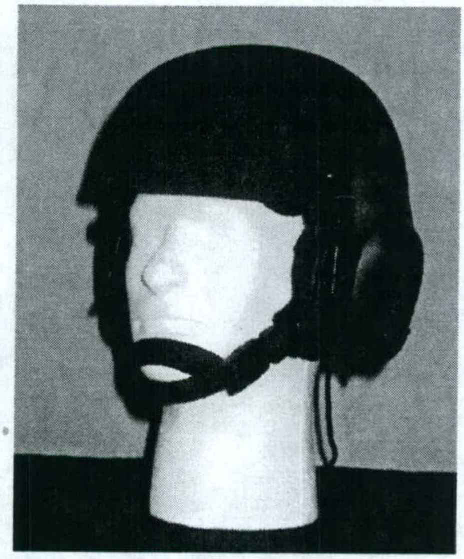


Figure 18: Transparent Hearing Platform, CGF/Gallet TC2001 with Sennheiser HD-205 headphones.

Additional headgear that was tested as aurally-occluding accessories include:

- Bacalava
- Dust Goggles
- Night-Vision System (PVS-14)
- Semi-Permeable Membrane (SPM) chem.-bio fabric
- JSLIST/XM-45 chem-bio mask and hood

4.2.2 Active Electronics Implementation

All non-commercial physical prototypes employ active electronics to control the signal path to the listener. Common among all prototypes is the use of the Panasonic WM-61 electret microphone capsule, see Figure 19. The capsule's specifications are given in Table 4: Specifications for the Panasonic WM-61 electret microphone capsule. Two hundred (200) microphone capsules were purchased and tested for linearity, gain, and spectral response. The capsules were then grouped by their least-squared differences in gain and spectral response. All microphone pairs employed in this project were as well-matched as any available.

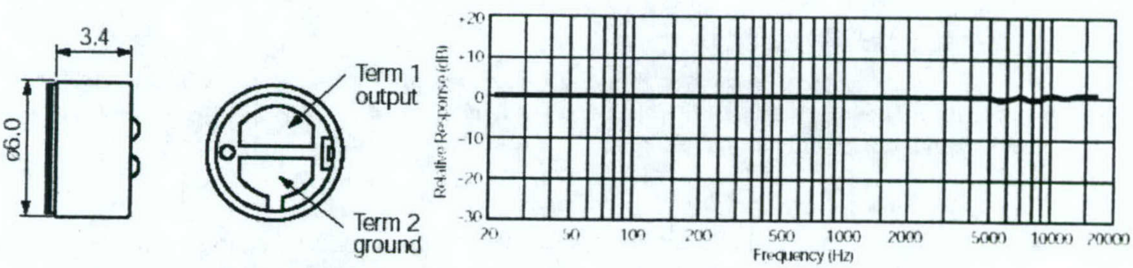


Figure 19: Panasonic WM-61 electret microphone capsule physical and electrical characteristics.

Table 4: Specifications for the Panasonic WM-61 electret microphone capsule.

Specification	Typical Value
Sensitivity	–35±4dB (0db = 1V/pa, 1kHz)
Impedance	Less than 2.2 k½
Directivity	Omnidirectional
Frequency	20–20,000 Hz
Max. operation	10V
Standard operation	2V
Current consumption	Max. 0.5 mA
Sensitivity reduction	Within –3 dB at 1.5V
S/N ratio	More than 62 dB

To power the multiple capsules required by most prototypes, a multi-channel electret interface board was created with balanced mic-level output. This multi-channel interface was compactly designed specifically for future use in helmet-mounted prototypes. The prototypes in the current study configured the interface to be belt-worn. The balanced lines allow long cables back to professional analog and digital audio equipment and support tethered roaming with worn prototypes about a 20-meter radius. Future wearable prototypes can place the interface in the helmet and processing electronics as vest-worn equipment. Final products would likely integrate all analog electronics in the helmet.

Prototypes that employed only a single binaural pair of microphones were connected through only a simple analog gain control before the signal was passed back into the Sennheiser HD-205 drivers. All prototypes were configured such that their signals could be digitally analyzed and/or filtered in real-time by the AuSIM3D digital audio signal processor, in parallel to the real-time listening. Future work with the binaural prototypes may involve digital filtering for equalization and automatic gain control in the signal control path before headphone delivery.

4.2.3 DSP Implementation

AuSIM's AuSIM3D digital audio signal processing system was used for all real-time filtering for systems employing digital algorithms. While all of the filters employed in the current work were non-adaptive, the AuSIM3D system is specifically designed to apply dynamically changing filters, which may be attempted in future work.

Filters were designed off-line and loaded into the AuSIM3D engine via the acoustic head map facility. All computation was 32-bit IEEE floating point. The most common sample rate was 48 kHz. Some work was performed at 96 kHz and is so noted in the results presentation. The total system latency was 3 buffers of 64 samples, equating to 4 msecs at 48 kHz, and verified by measurement. An Application-Specific Integrated Circuit (ASIC) implementation of these algorithms for a final product could perform with sub-millisecond latency.

4.2.4 East Coast Laboratory Approaches

This section describes the approaches implemented by the East Coast Laboratory, including the optimization method selection, reference system selection, and approach design.

4.2.4.1 Optimization Method Selection

After analysis, the LSM solution $\{W_m(f)\}_{\mathrm{lsm}}$ provides a more accurate match of the desired pinna cue magnitude response than the LSE solution, which can lead to improved acoustic transparency[7]. This improved performance comes at the cost of an optimization problem with no closed form solution, however, which can only be solved by means of a numerical search algorithm. This problem is time-

[7] The Least-Squared Error (LSE) and Least-Squared Magnitude (LSM) optimization methods were described in section 3.2.2.5 above.

consuming to solve, and experience indicates that reliable solutions are obtained only for simulated pinnae systems with 4 microphones or fewer per pinna.

Figure 20 shows an example of the left-ear behavior of these general optimization techniques. Specifically, it compares magnitude response as a function of frequency and elevation angle for (a) the desired HRTF (from a KEMAR manikin), (b) the LSE simulated pinnae system, and (c) the LSM simulated pinnae system for sources arriving from azimuths of 0 degrees. The LSE simulated pinnae system of Figure 20 (b) has difficulty matching the desired elevation-dependent spectral notch pattern, which is due largely to the overly-broad error optimization that seeks to match phase as well as magnitude information. The LSM simulated pinnae system of Figure 20 (c), on the other hand, matches the desired notch pattern comparatively well, which is due largely to the magnitude-only response optimization.

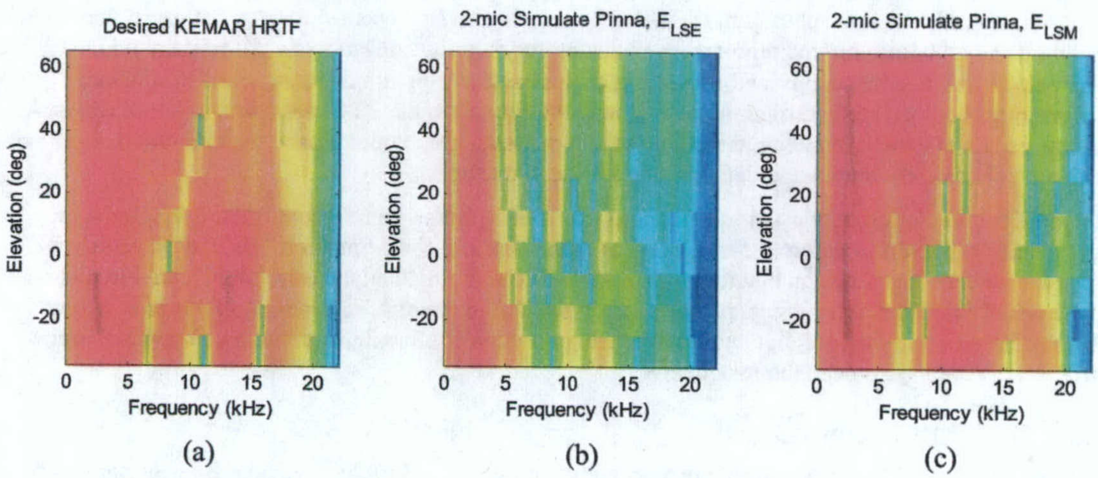


Figure 20: Left-ear simulated-pinna system example. Panels show system magnitude response for sources arriving from 0 degrees azimuth and -30 to 60 degrees elevation. Plots are images of response magnitude as a function of frequency (horizontal) and elevation (vertical). Larger magnitude = red, intermediate magnitudes = yellow, and smaller magnitude = blue. (a) Desired (KEMAR) HRTF. (b) LSE simulated pinnae. (c) LSM simulated pinnae.

Figure 21 shows the left-ear behavior of both the desired and the delay-and-sum simulated-pinnae system as a function of frequency and elevation angle for sources arriving from azimuths of 0 degrees. The elevation-dependent notch for the delay-and-sum system in Figure 21 (b) is strongly evident and follows the notch of the desired response shown in Figure 21 (a). Note that the notch is more strongly evident for the delay-and-sum simulated-pinna system than it is for the LSM system shown in Figure 20 (c). This is expected, since the delay-and-sum system has the primary goal of reproducing the notch, while the LSM system also tries to preserve other features of the desired HRTF.

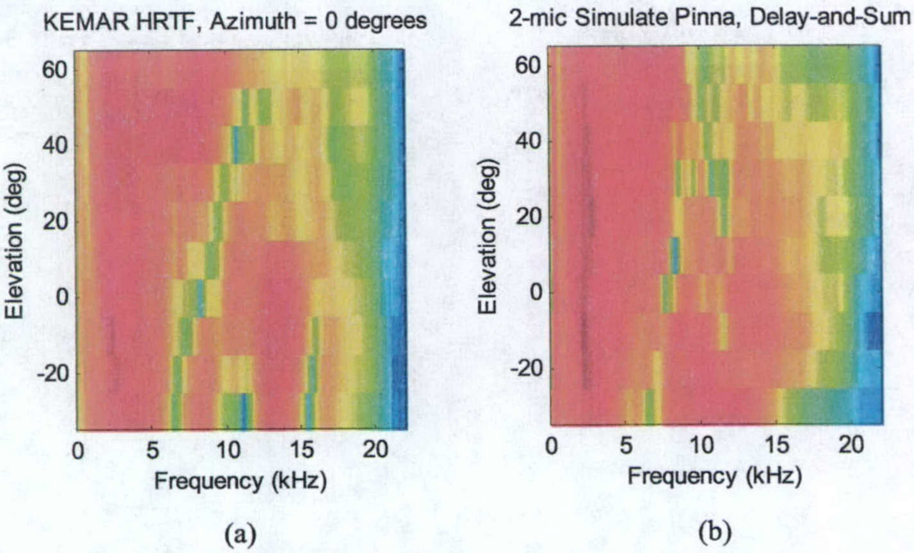


Figure 21: Left-ear delay-and-sum simulated-pinna system example. Panels show system magnitude response for sources arriving from 0 degrees azimuth and -30 to 60 degrees elevation. Plots are images of response magnitude as a function of frequency (horizontal) and elevation (vertical). Larger magnitude = red, intermediate magnitudes = yellow, and smaller magnitude = blue. (a) Desired (KEMAR) HRTF. (b) Delay-and-Sum Simulated Pinna.

Figure 22 shows the left-ear behavior of the 14-mic general microphone array system as a function of frequency and elevation angle for sources arriving from azimuths of 0 degrees. The most notable property of the general microphone array performance in Figure 22 (b) is that it does not preserve the desired HRTF spectral cues above 4kHz very well. The reason for this poor performance lies in the overly-broad least-squared error design metric used to create the system. As stated above, the LSE metric attempts to match both the magnitude and phase of the desired HRTF. With many microphones over which to optimize, the system can become overly influenced by the desired phase information. Currently, there is no alternative general microphone array system design technique that yields consistent reliable results, and so this research considers only LSE-based general microphone array systems.

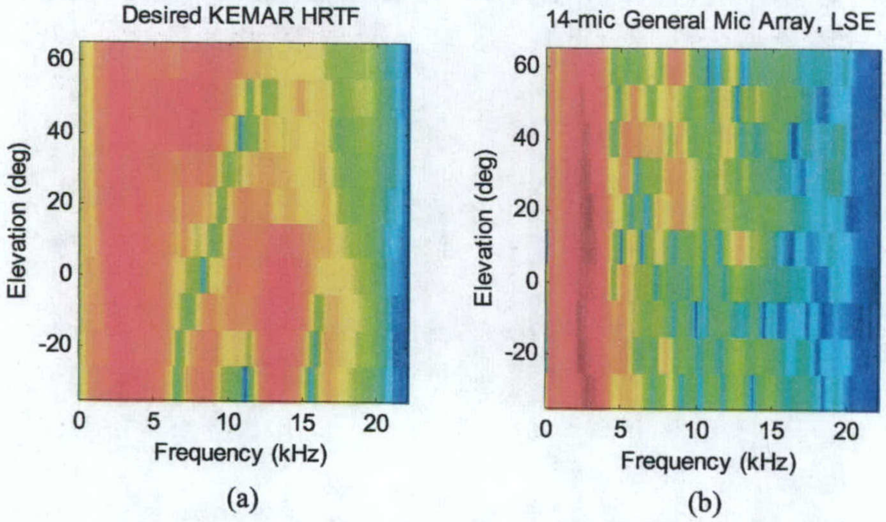


Figure 22: Left-ear 14-mic general microphone array system example. Panels show system magnitude response for sources arriving from 0 degrees azimuth and -30 to 60 degrees elevation. Plots are images of response magnitude as a function of frequency (horizontal) and elevation (vertical). Larger magnitude = red, intermediate magnitudes = yellow, and smaller magnitude = blue. (a) Desired (KEMAR) HRTF. (b) 14-mic general microphone array.

4.2.4.2 Physical Implementations

Table 5 summarizes the test systems used to evaluate the hidden concha, simulated pinnae and general microphone array approaches. This test set consisted of three reference systems, one hidden concha system, 7 simulated-pinnae systems, and 2 general microphone array systems. All artificial systems (i.e., not based upon the open ear) in this set, with the exception of the hidden concha system, were evaluated using both the physical evaluation metrics and the human localization test.

Table 5: Reference, hidden concha, simulated-pinnae, and general microphone array test systems.

	System	Mics Left/Right	Desired HRTF	Processing
REFERENCE	OE	N/A	N/A	User's open ears are tested.
	OE-H	N/A	N/A	User's open ears while wearing helmet.
	PEL	1 mic	N/A	PELTOR AGC muff.
HIDDEN CONCHA	HC_KE	1 mic in each 'ear canal'	KEMAR	1 mic/ear. Hidden concha system with KEMAR-based conchae.
SIMULATED PINNAE	SP1a_LSM SP1b_LSM	a = 1L/1R b = 3L/3R	KEMAR, Custom	1 omni/ear. Single omni mic near each ear. Filter applied to obtain general shape of desired HRTF.
	SP2a_LSM SP2b_LSM	a = 1,2L/1,2R b = 3,4L/3,4R	KEMAR, Custom	2 omnis/ear. Filters minimize LS dB magnitude error between output DTF and desired HRTF.
	SP4_LSM	1-4L/1-4R	KEMAR, Custom	4 omnis/ear. Filters minimize LS dB magnitude error between output DTF and desired HRTF.
	SP2a_DEL SP2b_DEL	a = 1,2L/1,2R b = 3,4L/3,4R	KEMAR, Custom	2 omnis/ear. Upper omni delayed and added to lower omni to create elevation-dependent notch. Delay chosen to approximate desired HRTF notch. Filter applied to obtain general shape of desired HRTF.
GENERAL MIC ARRAY	GA8C_LSE	1,2,5,7L and 1,2,5,7R for both ears	KEMAR, Custom	8 omnis/ear. General filters minimize LS complex distance error (magnitude and phase) between output DTF and desired HRTF.
	GA14C_ LSE	1-7L and 1-7R for both ears	KEMAR, Custom	14 omnis/ear. General filters minimize LS complex distance error (magnitude and phase) between output DTF and desired HRTF.

Figure 25 (PEL), Figure 26 (HC), Figure 28 (SP), and Figure 29 (GA) show photographs of the test systems as mounted on the KEMAR manikin. All test systems, with the exception of OE, also required the test subject to wear the MICH helmet with cutaway ear areas in addition to the hearing protective muff. Hearing protection for all non-reference systems was provided by Sennheiser HD-205 muffs, and sound for these systems was presented to the wearer through receivers located in these muffs.

The simulated-pinnae and general microphone array systems used specific subsets of the 14 microphones mounted on the muffs and helmet of the test system as shown in Figure 23 and Figure 24. These 14 microphones were divided into left and right sets of 7. Within each set, microphones 1-4 were arranged towards the front of the muff located approximately at the corners of a 1cm square. Microphones 5-7 were arranged across the appropriate side of the helmet. Several sets of filters were designed for each of these systems (according to the methods described in Section 3.2.2.5) in order to create system DTFs that matched the KEMAR manikin HRTF as well as the individual test-subject HRTFs. All systems were designed using a spatial error weighting function -- $w_{\mathrm{lse}}(f,\theta,\phi)$ or $w_{\mathrm{lsm}}(f,\theta,\phi)$ -- that weighted the elevation-dependent spectral notches in the desired HRTF five times more heavily than other spectral features.

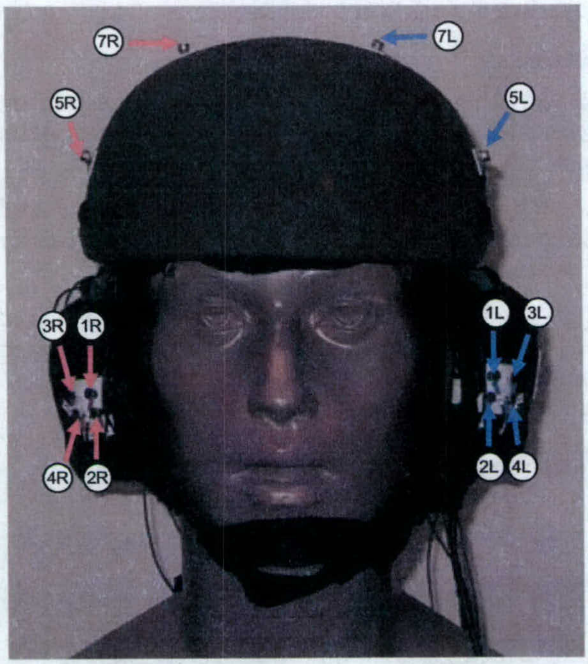


Figure 23: Front view of fourteen-microphone array apparatus used for simulated-pinnae and general microphone array test systems.

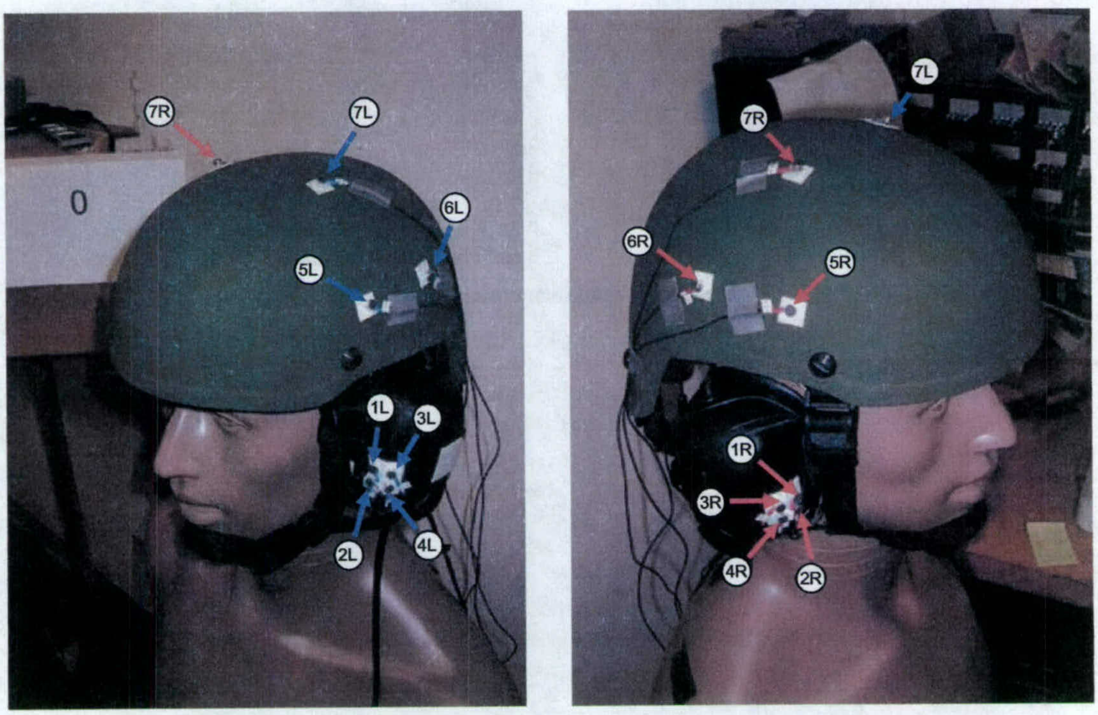


Figure 24: Side views of fourteen-microphone array apparatus used for simulated-pinnae and general microphone array test systems.

The detailed description of these test systems is as follows:

4.2.4.3 Reference Systems

- Open Ear **(OE)**: This 'system' consisted of using the test subjects' own natural, unaltered hearing. OE represented the baseline in performance that all artificial acoustic-transparency test systems were designed to attain.
- Open Ear with Helmet **(OE-H)**: This system consisted of using the test subjects' own natural hearing while wearing the MICH helmet with cutaway ear areas. OE-H represented situations involving soldiers wearing head, but not ear, protection.
- Peltor AGC Muff **(PEL)**: This system, shown in Figure 25 (a), consisted of the commercially-available Peltor COMTAC hear-through muffs, similar to circumaural communication devices currently used by the military. These muffs have automatic gain control (AGC) that operate independently in the two muffs.

Figure 25: Peltor COM-TAC protective hear-thru muff, in detail (left) and as mounted on KEMAR (right).

4.2.4.4 Hidden Concha Systems

- KEMAR-based hidden concha (**HC-KE**): This system, shown in Figure 26, consisted of KEMAR-molded hidden conchae recessed within the Sennheiser hearing protective muff and fed through to the respective ear.

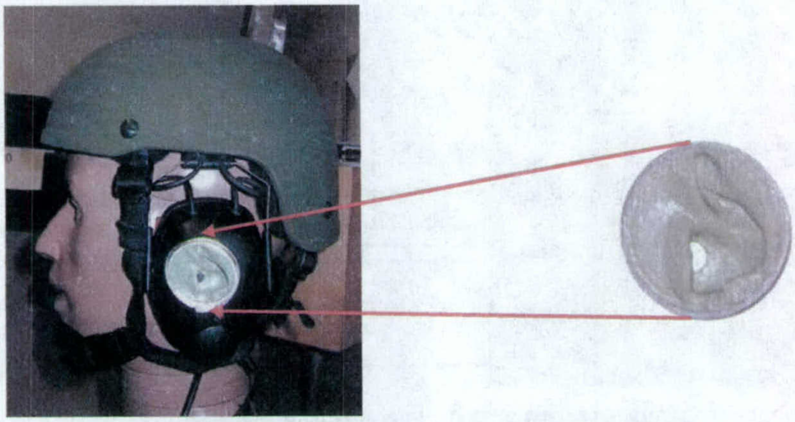


Figure 26: Hidden-concha system showing molded human-like concha modeled after KEMAR embedded into a Sennheiser HD205 protective hearing muff.

Figure 27 shows an example of the behavior of a hidden concha system. Specifically, it compares magnitude response as a function of frequency and elevation angle for (a) the desired HRTF (from a KEMAR manikin) and (b) the KEMAR-based hidden concha system for sources arriving from azimuths of 0 degrees. As indicated in Figure 27 (a), the most evident elevation-dependent spectral feature of the desired HRTF is a notch in frequency that changes with elevation – starting at about 6 kHz at low elevations and increasing to about 11 kHz at higher elevations. The hidden concha system follows the broad characteristics of the desired HRTF, although the elevation-dependent notch is not as evident.

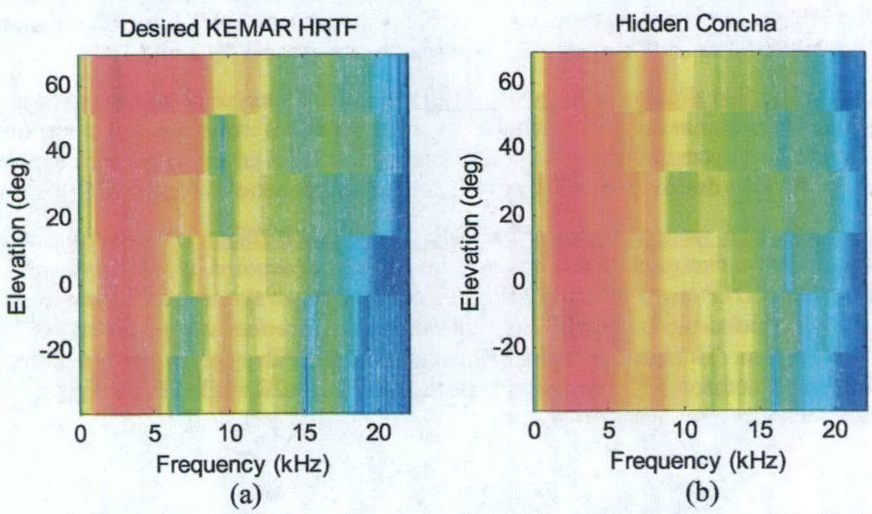


Figure 27: Left-ear hidden concha system example results. Panels show system magnitude response for sources arriving from 0 degrees azimuth and -30 to 60 degrees elevation. The plots are images of response magnitude as a function of frequency (horizontal) and elevation (vertical). Larger magnitude = red, intermediate magnitudes = yellow, and smaller magnitude = blue. (a) Desired (KEMAR) HRTF. (b) Hidden Concha.

4.2.4.5 Simulated-Pinnae Systems

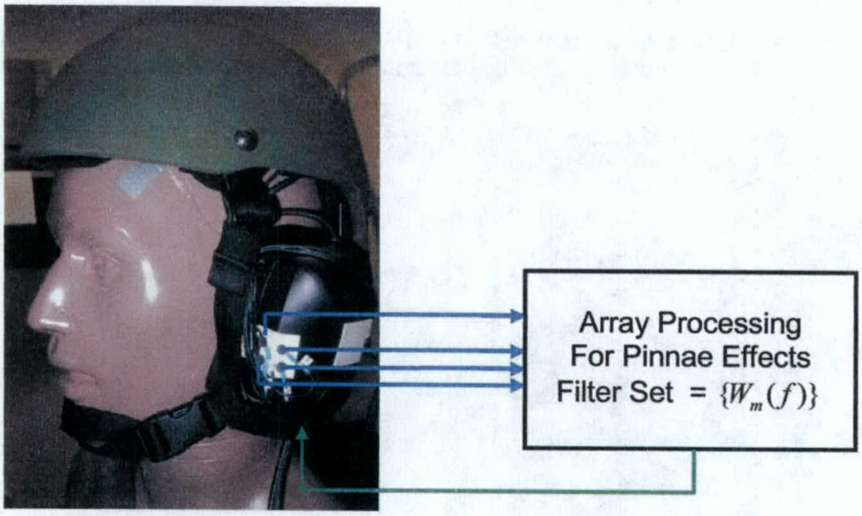


Figure 28: Simulated pinna prototype showing microphone placement.

- 1-microphone LSM simulated-pinnae (**SP1a-LSM** and **SP1b-LSM**): Each of these two systems used one microphone per ear to create simulated-pinnae outputs. The system filters were designed using the LSM criterion to match the average system DTF to the desired HRTF. The difference in the SP1a versus the SP1b systems was in the microphone placement: SP1a used microphones 1L and 1R while SP1b used the more widely spaced microphones 3L and 3R.
- 2-microphone LSM simulated-pinnae (**SP2a-LSM** and **SP2b-LSM**): Each of these two systems used two microphones per ear to create simulated-pinnae output. At each ear, the microphones were oriented vertically and separated by 1cm. The system filters were designed using the LSM criterion to match the average system DTF to the desired HRTF. The difference in the SP2a versus the SP2b systems was again in the microphone placement: SP2a used microphones 1,2L and 1,2R while SP2b used the more widely spaced microphones 3,4L and 3,4R.
- 4-microphone LSM simulated-pinnae (**SP4-LSM**): This system used four microphones per ear to create simulated-pinnae output. At each ear, the microphones were arranged at the corners of a 1cm square. The system filters were designed using the LSM criterion to match the average system DTF to the desired HRTF. The system used microphones 1-4L and 1-4R.
- 2-microphone DEL simulated-pinnae (**SP2a-DEL** and **SP2b-DEL**): The system filters for these two systems were designed by the delay-and-sum method (denoted by 'DEL') to match the average system DTF to the desired HRTF. Each of these two systems used two microphones per ear to create simulated-pinnae output. At each ear, the microphones were oriented vertically and separated by 1cm. The difference in the SP2a versus the SP2b systems was again in the microphone placement: SP2a used microphones 1,2L and 1,2R while SP2b used the more widely spaced microphones 3,4L and 3,4R.

4.2.4.6 General Microphone Array Systems

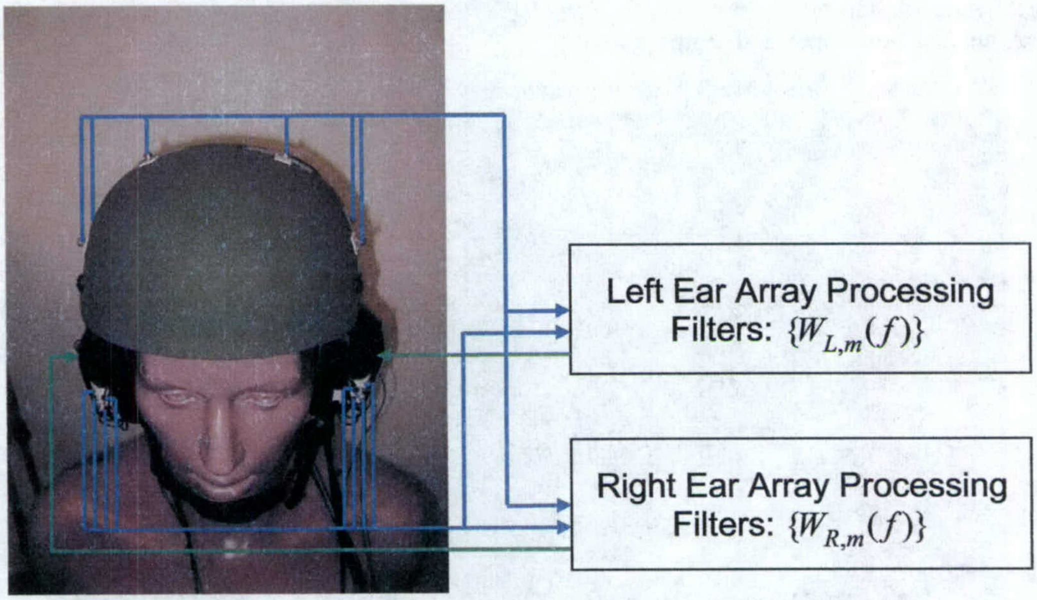


Figure 29: General microphone array prototype showing microphone placement.

- <u>8-microphone LSE general microphone array (**GA8C-LSE**):</u> This system used eight microphones in common to create the system output. Specifically, it used microphones 1,2,5,7L and 1,2,5,7R. The system filters were designed using microphones 1L and 1R to create the left and right binaural low-pass reference signals and the LSE criterion to match the average system DTF to the desired HRTF for high-pass signals.
- <u>14-microphone LSE general microphone array (**GA14C-LSE**):</u> This system used all fourteen microphones in common to create the system output. The system filters were designed using microphones 1L and 1R to create the left and right binaural low-pass reference signals and the LSE criterion to match the average system DTF to the desired HRTF for high-pass signals.

4.2.5 West Coast Laboratory Approaches

This section describes the approaches implemented by the West Coast Laboratory, including the optimization method selection and approach design.

4.2.5.1 32-channel Helmet/Muff Microphone Array

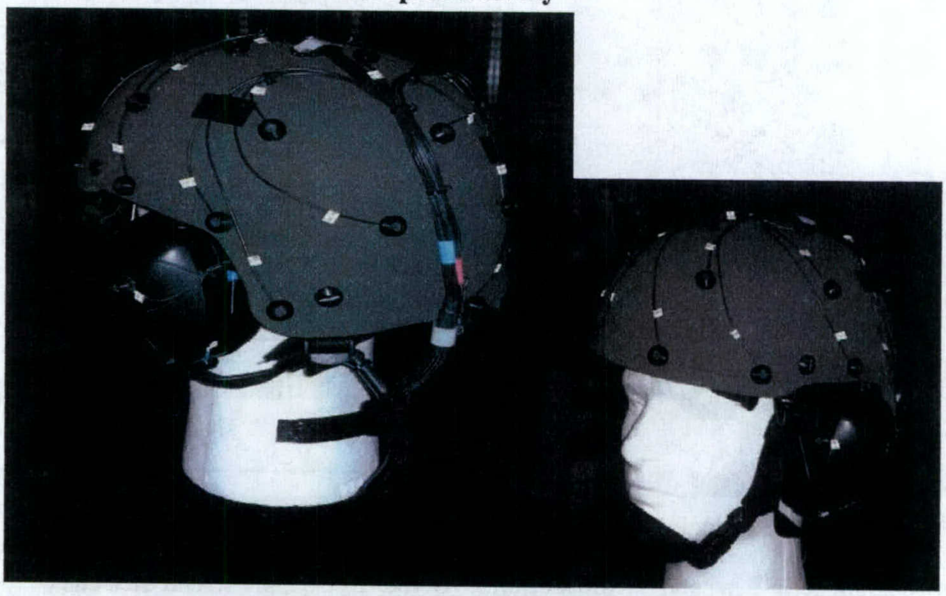


Figure 30: 32-channel system prototype, showing 4 microphones located on each muff, 4 microphones on the sagittal plane of the helmet, and 10 pairs of microphones reflected on each side of the helmet.

Overview

The first system studied consisted of 32 microphones distributed uniformly around the helmet and muffs. In the arrangement, 8 microphones were placed on the two muffs and the remaining 24 distributed around the helmet. There were several goals for the helmet array.

1) **Microphone Location Sensitivity on HeadGear.** By distributing a large array of microphones across the headgear platform, simultaneous response at several locations to different directional simuli could be observed.

2) **Microphone Directivity Due to HeadGear Shadowing.** Tightly related to the positional sensitivity across the surface of the headgear are the directional characteristics of each position. The optimal type of array processing is dependent on directional sensitivity or more accurately the directional acuity for each microphone.

3) **Study of Waveform Flow Across the HeadGear Surface.** By employing a large array, the temporal flow of sound pressure gradients across the surface can be observed, headgear accessories can be applied and the acoustic disturbance can be observed. Potentially, the array processing filters can be altered to automatically account for these accessories.

4) **Study of a Baseline Reference Prototype Employing Direct HRTF Filtering.** The simple approach described in section 3.2.2.8 assumes exclusive-directivity of each microphone.

5) **Basis for Filter and Location Optimization.** Because the microphones have overlapping directivity responses, the primary goal of this prototype was to search for an optimal set of filters that would give a better result than straight HRTF filters.

Once a good system was developed employing the 32-channels, a search on all possible subsets and exploration of the performance vs. channel-count tradeoffs could be performed.

Microphone Mounting Design

In order to maximize the directivity of the mounted microphones for the 32-channel array, a variety of possible microphone mounts were explored. The microphone mount prototypes were a variety of plastic and rubber cylinders with a hole in the middle to house the microphone. The mount cylinders had different diameters and heights, and some had beveled edges. To test the directivity of the mounts, these were placed on a flat baffle, and the response was measured at different wave-front incidence angles.

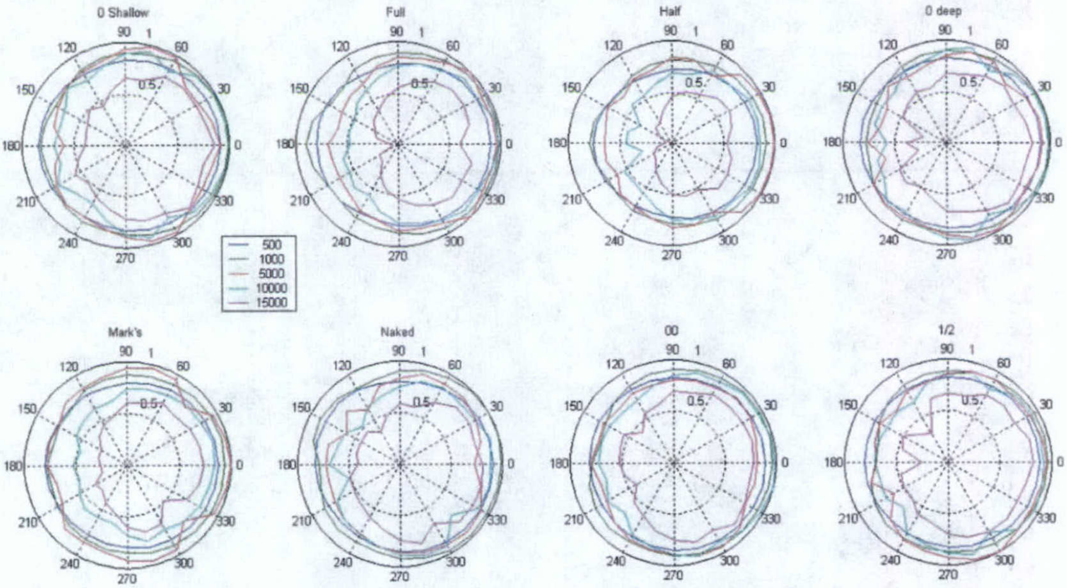


Figure 31: Directivity measurements of selected microphone mount prototypes. The "0-deep" prototype, shown in the upper right, was selected for 32-channel system.

Based on these directivity measurements, the mount labeled as "0 deep" was selected as the best mount. This mount resulted in a fairly flat frequency response with an acceptable directivity pattern, especially at high frequency.

Microphone Distribution

After several iterations, a microphone layout was selected consisting of 4 microphones distributed on each headphone muff, 10 microphones on each side of the helmet, and 4 unpaired microphones on the medial (sagittal) plane. This layout, shown in Figure 30, is a compromise between optimal geometric distribution featuring equi-spacing and the constraints of the helmet and muff surface availability. Despite the compromise, this layout affords a study of permuted combinations of 28 pairs of microphone locations. The 4 microphones on the sagittal plane provide signal data without interaural differences. The resulting native microphone directivity of this distribution is depicted in the triple panes of Figure 32 below.

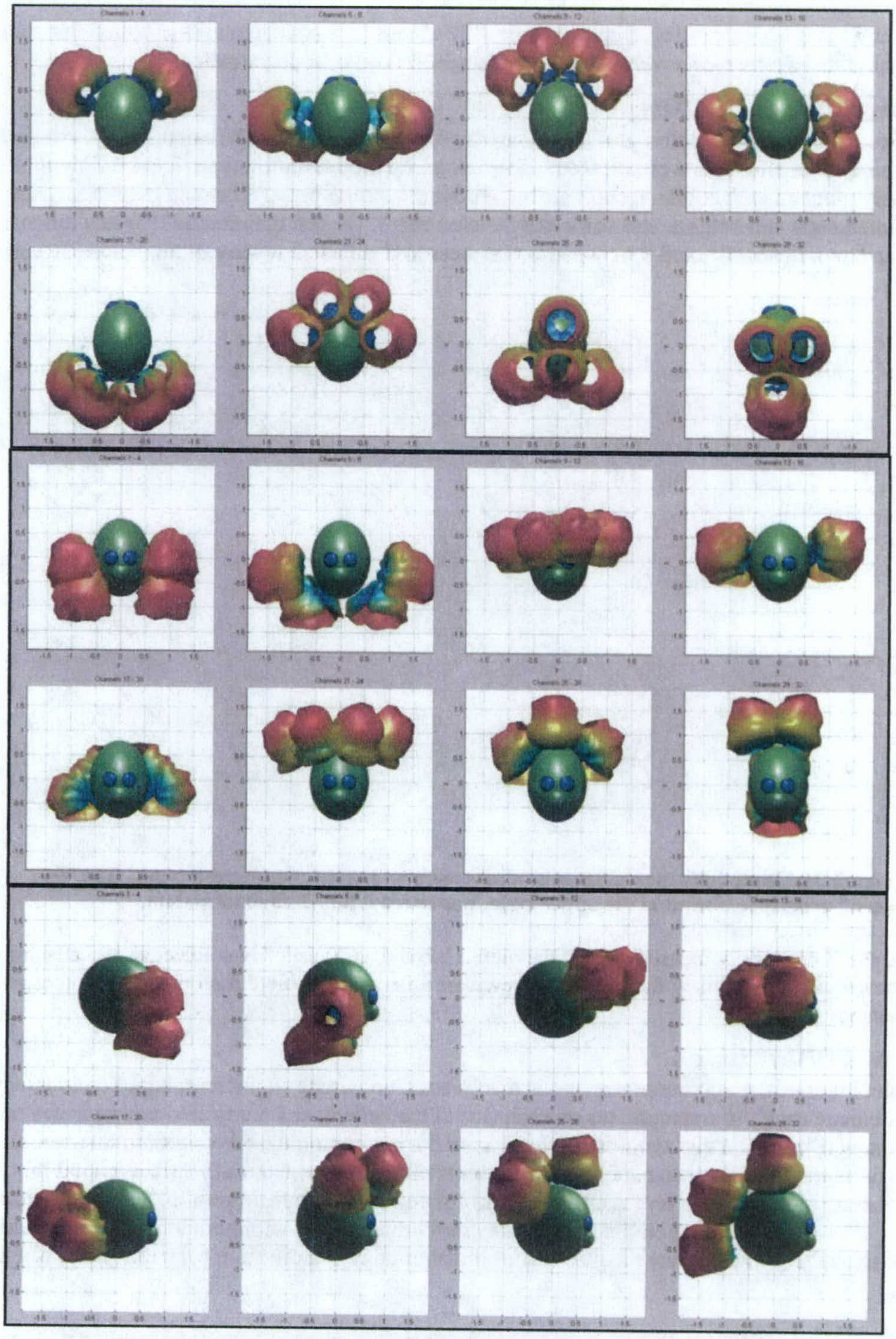


Figure 32: Polar surface plots of full spectrum directivity of microphones. Color and polar magnitude is proportional to the magnitude of the microphone response in the given polar direction. Plot origin corresponds to microphone's physical location. Data captured with the AuSIM HeadZap measurement system.

Direct HRTF

For the direct HRTF filtering approach, the physical location and orientation of each of the microphones was measured first. To do the measurement, a 6 degree-of-freedom Polhemus tracker sensor was placed above each of the microphones and the location was sampled. Next, a virtual environment was created with a virtual sound source located at the same location as the measured microphone location. This resulted in each microphone being filtered by the HRTF in that direction.

DTF to HRTF Filter Optimization

For any of the microphone array systems, the processing diagram is fundamentally the same where each microphone-ear pair has a filter applied, as shown in Figure 33. For the direct HRTF approach, the filter that is applied is an actual HRTF. However, better results may be possible by using numerical optimization to find the "optimal" set of filters that will generate an output that is close to the ideal output of the system.

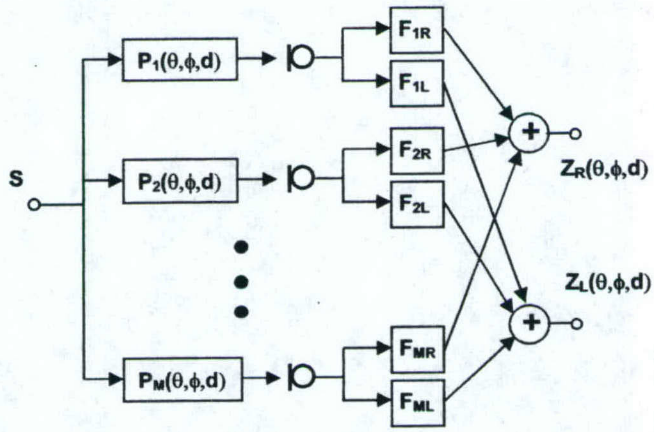


Figure 33: Transparent Hearing schematic block diagram showing processing blocks.

The most straightforward method of filter optimization is to use the closed-form complex frequency domain least-squared error optimization, as defined by Equation (5) above. When applied to the 32-channel array, the resulting frequency responses had reasonably ideal performance; however, when trying to convert the responses to the time domain for filtering, the resultant impulse responses had fairly constant energy over time, causing significant distortion due to circular convolution effects. The length of the impulse response can be controlled by the addition of a regularization parameter which adds a small amount of leakage to the optimization. By keeping the length of the impulse response under control, it is possible to minimize the errors introduced by circular convolution [74][131].

DTF to Beams to HRTF Optimization

The beam-formed optimization is a variation on the previous filter optimization of the 32-channel array, which tries to leverage the benefits of the Direct HRTF method. In the Direct HRTF method, each microphone was assigned a different HRTF filter, in the hope that the microphone would capture any sound in that direction, and present it as coming from that direction. However, since the microphones have a fairly wide directivity pattern, there is a large amount of overlap between the responses of the microphones. This method tries to find a set of signals or "beams" that are spatially independent and can be subsequently filtered by an HRTF in that direction to give a good result.

The final system diagram for this method will ultimately be the same as for the previous 32-channel filter optimization, as shown in Figure 33. However, conceptually there will be two sequential banks of FIR filters. The first bank will be the set of filters which will optimally generate the set of desired beams.

The second set will be a set of HRTF filters. These two filter banks can be pre-convolved to give a set of filters that fit into the original system diagram.

The filter optimization problem then becomes to search for the set of filters that will best match a set of ideal beams. An ideal beam is defined as giving a flat frequency response in the direction of the beam, and no frequency response in any other direction. So, the same optimization techniques can be used to perform the optimization, but rather than searching for the filters that match the ideal HRTF, the object of the search is the filters that match the ideal beams.

The ideal beams formed look very promising. A given beam can select one direction above all the others with about 15dB rejection, as shown in Figure 34. However, the beam-formed system result does not subjectively perform as well as optimizing directly to the HRTF filters.

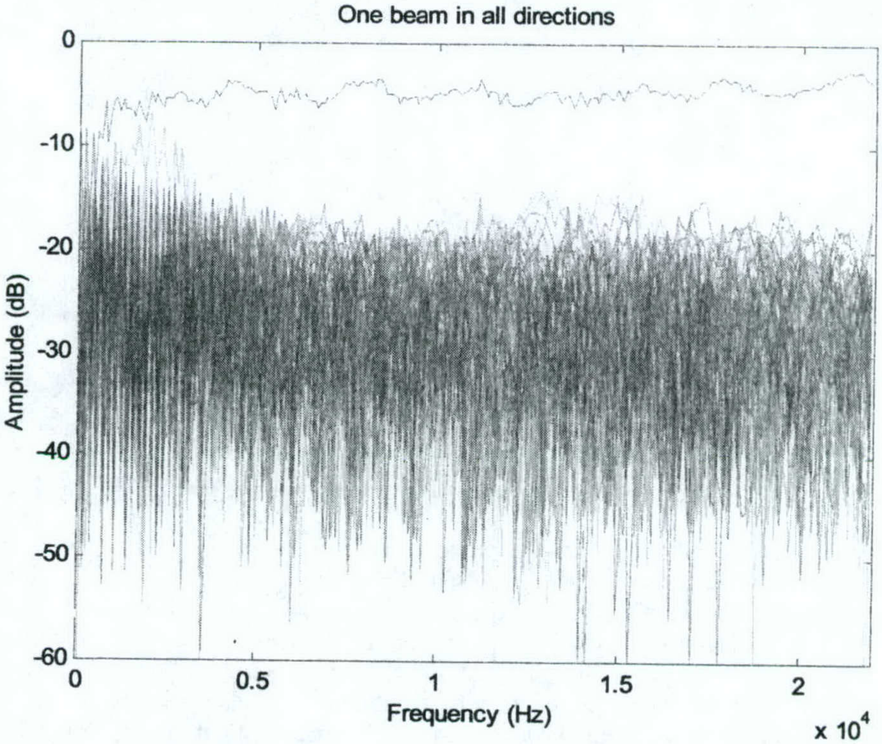


Figure 34: Magnitude plot for a selected signal with its optimized direction (isolated line) against all other directions (grouped lines) with the beam-forming algorithm applied to optimize directional isolation.

4.2.5.2 Muff-Mounted Pinnae

The implementation of pinnae mounted on hearing-protector muffs is an attempt to explore the approach of binaural sensing through human-replica pinnae discussed in Section 3.2.2.2 above. This approach calls for a pair of microphones on opposing sides of the listener's head, each in the canal of a human life-like pinna. This approach intends to follow the success of binaural "dummy heads" for delivering life-like spatial audio cues.

The implementation in the present study began by slicing the most feasible depth away from the base Sennheiser HD205 headphone muff enclosure to create a flat mounting surface. A flat baffle was then fabricated to fit nicely around the open back of the muff. A ring was fabricated for the next layer to serve two purposes: (a) the inside edge retains the rubber-like pinna material, and (b) the outside edge retains an optional open-cell foam wind-screen and cover. A canal of the proper anthropometric diameter is machined through the baffle with a microphone mounted at the bottom. For implementation in this study, pinnae were considered from those custom-molded at Wright-Patterson AFB, generically-molded for KEMAR by Knowles, and generically-molded for the KU-100 by Neumann. The Neumann's were judged the most suitable for this prototype and employed. The result is shown in Figure 35.

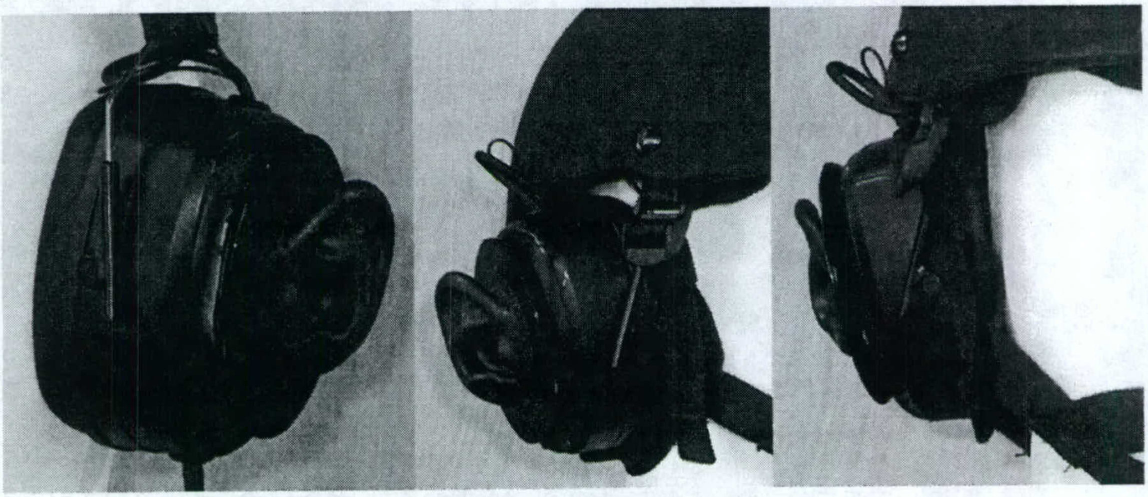


Figure 35: Multiple-views of the muff-mounted human-replica pinna system. The extra flange just outside the outer edge of the muff is the constraining ring for the optional windscreen, not shown.

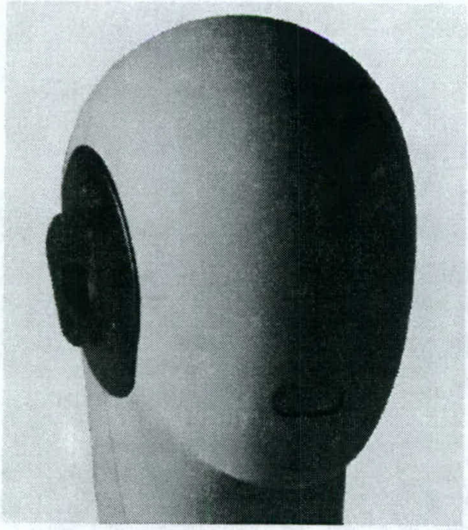
Both with and without the windscreen attached, this approach generates the physically widest hear-thru device implementation of those tested in this study. The space constraints are created by a layering of systems. The first layer is a circumaural seal designed to be larger than the listener's pinna for maximum comfort and mid-frequency protection. The second layer is the acoustic transducing driver for the phonic display. The driver employed is a relatively small diaphragm (2 cm diameter) with less than a centimeter in total thickness. The third layer is sound isolation which includes a thick wall of plastic and acoustic baffling material. The fourth layer is the binaural microphone capsule housing. The fifth layer is the simulated ear canal, with an approximate depth of 8 mm. And finally the sixth layer is the molded pinna including the full depth of the concha cavity.

4.2.5.3 Mechanically-Modeled Pinnae

The mechanically-modeled pinna approach is an attempt to find a simplified 3D geometry that can physically encode sound waves with direction specific characteristics. Several distinct explorations were made within this approach. The implementation began with a study of existing simplified pinna. Examples include the Head Acoustics instrumented mannequin shown in Figure 36. Even though this study did not obtain acoustic data from the Head Acoustics device, it served as an inspiration for design. A future study could investigate the variety of animal pinna which yields excellent directional sound perception.

Other relevant previous work includes the studies by Shaw on replicable concha shapes. This geometry afforded Shaw a finite set of geometric variables for study. For this project, a pinna was machined with reference to Shaw's geometry. The result is shown in

Figure 37. A full prototype hear-thru device utilizing Shaw's design was not developed to a wearable form, but the basic pinna was acoustically tested against a baffle to yield strong elevation and front-back spectral cues.



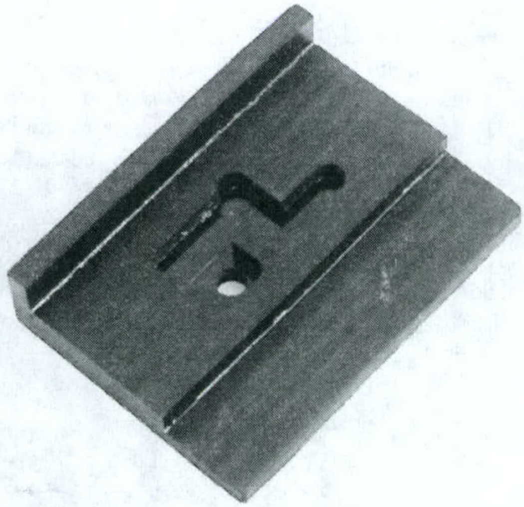


Figure 36 (left): Head Acoustics mannequin showing simplified pinna and concha shape.

Figure 37 (right): Machined pinna shape inspired by Shaw.

A wearable prototype device employing simplified pinna shape was developed using a similar technique to the muff-mounted pinnae system in the previous section. The implementation, as shown in Figure 38, began by removing the maximum depth from the Sennheiser HD205 headphone muff enclosure, to provide for a mounting surface with the tightest location to the underlying human ear. Approximately 1cm was removed from the stock HD205 enclosure, while still retaining the headphone's active and passive function. A flat baffle was then fabricated to fit nicely to the open back of the muff. A canal of the proper anthropometric diameter and a concha with simplified human characteristics were machined through the baffle with a microphone mounted at the bottom. The outer pinna consists of a wing-shaped baffle at an appropriate size and angle of human outer pinnae, but without the folds.

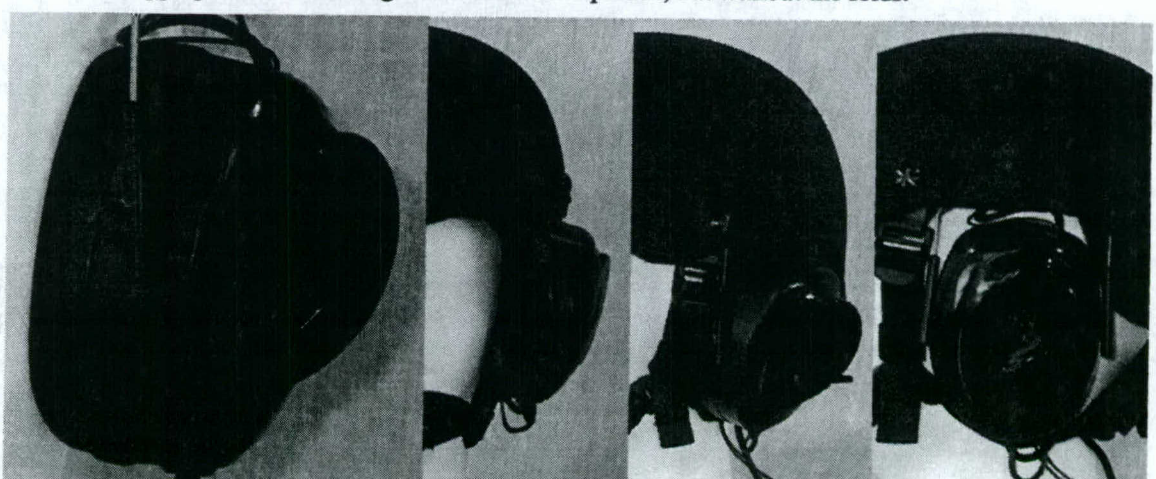


Figure 38: Hard Pinna mounted on muff for mechanically modeled pinna approach. The concha has human-replica features. The outer pinna is a much simplified baffle compared to a human equivalent. The ear canal depth is approximately 8 mm.

For implementations employing double protection and delivering the signals through an in-canal device, an additional 8 mm of depth can be removed.

Several ideas were explored which would utilize positions on headgear retaining physical characteristics of the sound signals and leverage local headgear features to give the signal direction-dependent shape. Several of these designs are depicted in the Design Guide in section 1. One design that was physically prototyped integrated a concha shape into two alternative inviting locations on the Scorpion R2 helmet design. These designs are shown below in Figure 39.

Figure 39: Human-modelled concha integrated into a helmet system. In this case, a cocha replica was integrated into alternative positions: on the muff and on the helmet at an approximately correct interaural distance.

4.2.5.4 Sound-field Microphone Apparatus

The basic idea of this technique was to use 4 omni-directional microphones to capture the first order spherical harmonics around the head. Once these were captured, they could be rendered to give a good representation of the directional sound field.

Since the microphone cannot be placed in the free field, a single four-channel microphone cannot capture all directions. Also, the Ambisonic reconstruction does a good job at reproducing the directional pressure, but does not preserve phase delay very well. Therefore, this method used two separate four-microphone arrays to create two sound-field microphones, one on each muff, as shown in Figure 40. In this way, the microphones could capture the directional pressure at each ear, and recreate the directional spectral magnitude cues, while the location of the two arrays captured reasonable representations of the ILD and ITD.

Figure 40: Helmet-integrated sound-field microphone approach. The sound-field microphone on each muff is a geometrically constrained array of 4 electret capsules. To support a total of 8 mic capsules, an 8-channel pre-amplifier is employed to interface the device to a digital processing.

To create the sound-field microphone from four omni-directional capsules, the capsules were placed with one at the origin and the other three along the x, y, and z coordinate axes, as depicted in Figure 41. Under this arrangement, the B-format components are conceptually: $W = z$, $X = x - o$, $Y = y - o$, and $Z = z - o$. However, in practice, a filter was placed on each of the inputs and optimized to yield the ideal B-format impulse responses.

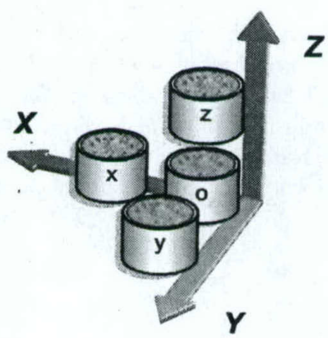


Figure 41: Compact sound-field microphone using 4 omni-directional capsules in a triad configuration.

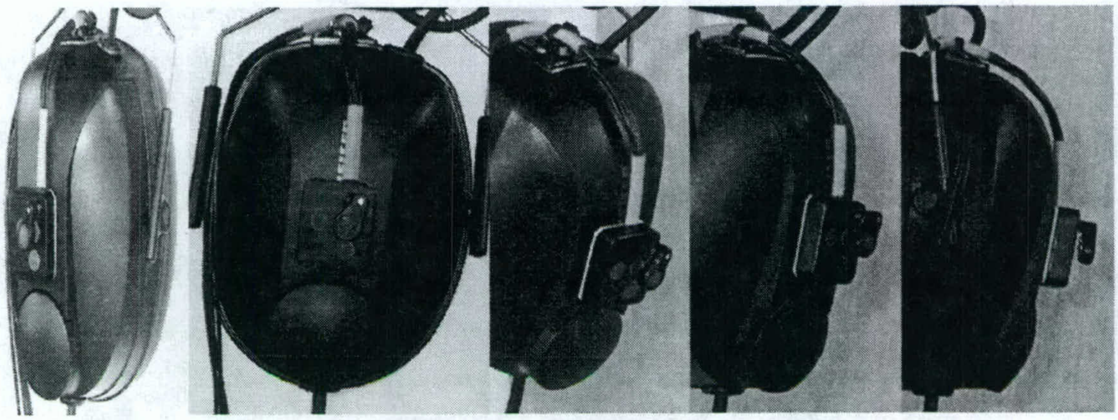


Figure 42: Compact sound-field microphone as mounted on Sennheiser HD205 muff. The array may be covered/protected with a hardwire cage and foam windscreen. The location is centered over the ear canal with respect to the sagittal plane, and thus on the interaural axis.

Even though the physical arrangement was inspired by an Ambisonic microphone, there are several options on how to render the microphones. The first option is to treat microphone arrangement as a general array, and try and optimize the maximum phase DTF of the array to the maximum phase HRTF, as was done for the 32-channel array. This did not give very good results because of the difficulties of optimizing for both phase and magnitude.

Since the microphones were in a cluster around the location where the time delay is fairly close to the desired time delay, it was possible to optimize to the minimum phase version of the HRTF. This allows for a much closer magnitude match, while maintaining a better time delay than is found by optimizing to both sound-field microphones collectively. For this method, the full maximum phase DTF was analyzed, and for each microphone array, the delays were removed to give a relative delay around the center of the array. These "almost minimum phase" responses were then used to find the optimum filters that would generate the correct minimum phase HRTF response for that ear.

The physical implementation can be improved by shaving down the depth of the muff as was done on the mechanical pinnae and human-pinnae prototypes. Additionally, the array can be better integrated into the shell of the muff. The implementation in the present work was simplified for emphasis on functional feasibility.

4.3 Evaluation

The implemented prototypes and commercially-produced systems under this study were evaluated using physical and perceptual methods. All systems were tested using a set of simple acoustic-measurement-based physical evaluation metrics. These metrics were chosen to highlight how well the various systems preserved some of the important HRTF localization features and serve as a preliminary estimate of system acoustic transparency. A subset of the systems was tested using a simple human-subject-based source localization test. This localization test process helped to assess the actual acoustic transparency of the test systems in a way that is not possible using acoustic measurements alone. Finally, selected COTS hear-through devices and earplugs were subjectively tested for sound quality, comfort, spatial cue retention and comfort.

4.3.1 Acoustic Testing

The evaluation of physical prototypes included empirical acoustic testing. The primary goal of such testing was to acoustically evaluate the performance of each approach, assess the differences between them, and determine their strengths and weaknesses. Additionally, acoustic testing was used in the prototype design phase for some approaches to ascertain various filter parameters, microphone placement, or geometry of mechanical elements. For multi-microphone systems, acoustic testing allowed direct comparison of the direction-dependent HRTFs expected from combination of earlier measurements and array processing.

Direction-dependent transfer functions were measured at a sample of source directions using AuSIM's HeadZap measurement system (see Appendix C). These measurements were compared to a reference response dataset by visually comparing plots and by using the metrics developed in this study as described below. System-generated noise, direction-independent system response, and system linearity were also characterized acoustically, as was the directional attenuation (or inversely "leakage") provided by the direct-path attenuator.

Instrumented mannequins (or "binaural dummy heads") and, in one case, a human head were used as test subjects for the acoustic testing. The available instrumented mannequins included KEMAR, Neumann KU-100, and Bruel & Kjaer's HATS, see Figure 43. All mannequins are equipped with internal microphones and detachable human-replica pinnae.

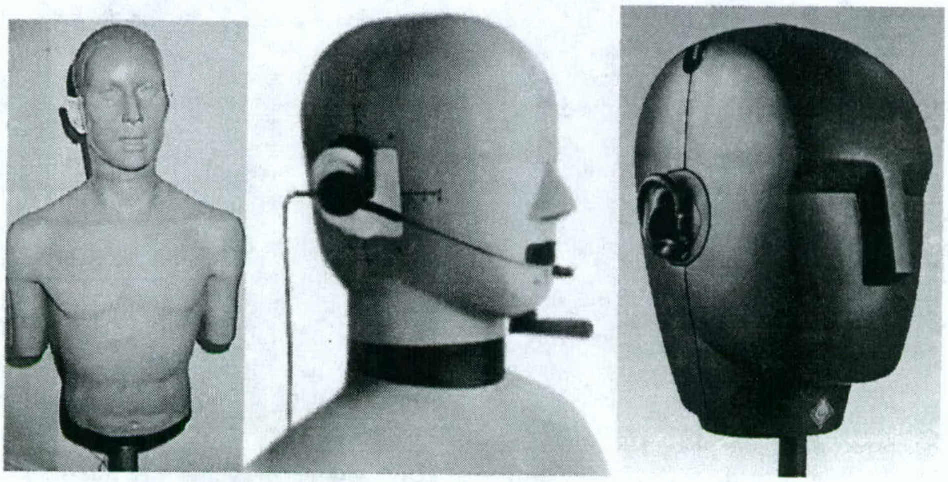


Figure 43: Instrumented test mannequins used in acoustic testing included (a) Knowles KEMAR, (b) B&K HATS, and (c) Neumann KU-100. All three feature replica human pinnae.

4.3.2 Evaluation Metrics

In order to assess the quality of transparent hearing systems in the design and test stages, metrics are needed for quantifying the deviation of a test system from a reference system. For the present study, the reference systems all included head-and-torso. Development of the metric was based on a survey of studies that have documented the physical cues for spatial localization.

4.3.2.1 Background

Metrics based on deviation from a reference (error) were developed and used in the design and optimization process needed to find filter weights for multiple-microphone systems, as described in section 3.2.2.5 above.

In one approach to developing metrics, a single grand metric could be formed. This metric might be a weighted sum of squared deviations between reference and test responses, summed over 1) all physical cues (monaural and interaural), 2) all directions[8], and 3) frequency. For example, one component in this

[8] Far-field will be assumed thus distance will not be a variable

sum would be interaural delay, which would presumably have relatively large weight at low frequencies (< 1500 Hz), but low weight at high frequencies. The problem with this single-metric approach is that it requires many weighting factors that are not well specified in the literature. Few studies have examined the relative potency of one type of localization cue versus another (e.g.,[143]) or of cue variations across frequency.

Another approach that was explored and implemented is a metric vector: multiple measures that are not explicitly combined. For example, there could be separate metrics for cues underlying left-right and elevation localization, or for ITD and ILD cues. The implementation of this approach is described below.

4.3.2.2 Error Metrics

Evaluation metrics for the test systems were divided into two classes: interaural difference metrics and spectral shape metrics. Interaural difference metrics measured the fidelity of the system interaural-cues and served as a measure of the system transparency for lateral azimuth-plane source localization. Spectral shape metrics measured the fidelity of the system spectral cues and served as a measure of the system transparency for source elevation.

Interaural-Difference Metrics

Two metrics were used to measure interaural cue fidelity: interaural time difference (ITD) error and interaural level difference (ILD) error. These metrics considered ITDs and ILDs in the 0-3.5 kHz frequency range, since these frequencies provided the most important binaural localization cues.

ITD error was defined as the RMS average ITD error:

$$\varepsilon_{ITD} = \sqrt{\sum_{\theta,\phi}\left[ITD_{\text{desired}}(\theta,\phi) - ITD_{\text{system}}(\theta,\phi)\right]^2 w_{ITD}(\theta,\phi)}\,, \qquad (16)$$

where $ITD_{\text{desired}}(\theta,\phi)$ and $ITD_{\text{system}}(\theta,\phi)$ were, respectively, the desired and system output ITDs as a function of location, and $w_{ITD}(\theta,\phi)$ was a location-dependent ITD weighting term. ITD was defined to be equal to the difference in left versus right channel group delay averaged over the 0-3.5 kHz range and then capped at a maximum delay value of $800\mu\text{sec}$:

$$\begin{aligned} A &= \sum_{f=0}^{3.5\text{ kHz}} GRD_{\text{left}}(f,\theta,\phi) - GRD_{\text{right}}(f,\theta,\phi), \\ ITD(\theta,\phi) &= \operatorname{sign}(A) * \min(800\mu\text{sec}, A). \end{aligned} \qquad (17)$$

The $800\mu\text{sec}$ cap results from the fact that ITDs in excess of approximately $800\mu\text{sec}$ are localized to the maximal ITD-induced source locations of $\pm 90^\circ$. The weighing term $w_{ITD}(\theta,\phi)$ was selected to weight centrally-located sources more heavily relative to lateral sources and was defined as

$$w_{ITD}(\theta,\phi) = 0.5 + \cos(\psi)\,, \qquad (18)$$

where ψ was the angle between the source location (θ,ϕ) and the mid-sagittal plane.

ILD error was defined in a more complicated manner. Specifically, it was equal to the weighted average of summed and transformed ILD error from 0-3.5 kHz:

$$\varepsilon_{ILD} = \sum_{\theta,\phi} w_{ILD}(\theta,\phi) \sum_{f=0}^{3.5\text{ kHz}} F\left[\left|ILD_{\text{desired}}(f,\theta,\phi) - ILD_{\text{system}}(f,\theta,\phi)\right|\right], \qquad (19)$$

where $ILD_{\text{desired}}(f,\theta,\phi)$ and $ILD_{\text{system}}(f,\theta,\phi)$ were the desired and system ILDs, respectively, as a function of location and frequency, $w_{ILD}(\theta,\phi)$ was a location-dependent ILD weighting term, and $F[\bullet]$ is an ILD error transformation term. ILD was defined to be equal to the dB difference in left versus right channel power capped at a maximum level difference of 10 dB:

$$\begin{aligned} B &= 20\log|H_{\text{left}}(f,\theta,\phi)| - 20\log|\alpha H_{\text{right}}(f,\theta,\phi)|, \\ ILD(f,\theta,\phi) &= \text{sign}(B) * \min(10\text{ dB}, |B|), \end{aligned} \tag{20}$$

where α was a broadband scale factor to equalize any volume difference between the left and right channels that was constant over all locations and, as such, could be accounted for relatively easily by the human perception system. The reason for the 10 dB cap was that ILDs in excess of 10 dB have roughly the same effect upon source localization. The weighing term $w_{ILD}(\theta,\phi)$ was selected to weight centrally-located sources more heavily relative to lateral sources and was actually equal to $w_{ITD}(\theta,\phi)$ as stated above. The ILD error transformation was chosen to weight ILD errors in excess of 5 dB ten times more heavily than ILD errors below 5dB:

$$\begin{aligned} F[\text{ILD\_Error}] &= \text{ILD\_Error} && \text{for ILD\_Error} < 5\text{dB}, \\ &= 10*\text{ILD\_Error} && \text{for ILD\_Error} \geq 5\text{dB}. \end{aligned} \tag{21}$$

This weighing reflects the fact that ILD errors must be quite large before they have a significant effect upon localization.

Spectral Cue Metrics

A single metric, the average magnitude response error, served as the measure of spectral cue fidelity. This metric covers the frequency range from 4.5 – 14 kHz, which is the range that covers the major spectral notches in the HRTF patterns that provide important elevation localization cues.

The average magnitude response error was defined as the average of the left and right channel weighted RMS dB magnitude response errors:

$$\varepsilon_{\text{mag}} = 0.5(\varepsilon_{\text{mag, L}} + \varepsilon_{\text{mag, R}}), \tag{22}$$

where

$$\varepsilon_{\text{mag, L/R}} = \sum_{f=4.5}^{14\text{ kHz}} \sum_{\theta,\phi} w_{\text{mag, L/R}}(f,\theta,\phi)\left[20\log|H_{\text{desired, L/R}}(f,\theta,\phi)| - 20\log|H_{\text{system, L/R}}(f,\theta,\phi)|\right]^2 .$$

In this definition, $H_{\text{desired, L/R}}(f,\theta,\phi)$ and $H_{\text{desired, L/R}}(f,\theta,\phi)$ were the left and right ear frequency- and location-dependent desired HRTF and system DTF, respectively, and $w_{\text{mag, L/R}}(f,\theta,\phi)$ were left and right ear frequency- and location-dependent weighting terms that emphasized important spectral features. Specifically, $w_{\text{mag, L/R}}(f,\theta,\phi)$ were based upon $H_{\text{desired, L/R}}(f,\theta,\phi)$ and weigh spectral notches roughly 5 times more heavily than the remaining spectral features.

4.3.3 Preliminary Subjective Testing

Prototypes were developed in two separate labs by the project team; thus, subjective testing was not duplicated for all prototypes. The west coast lab solicited informal subjective comments from many subjects. The east coast lab performed a localization test procedure on a limited number of subjects.

4.3.3.1 Localization Test Procedure

Sound localization performance by a small group of subjects was tested while they used various transparent hearing test systems, as well as their open ears. Subjects were tested individually in an office/shop room measuring 4.75m by 3.2m with a 3m ceiling. While keeping their eyes closed, subjects were asked to give azimuth and elevation estimates for a noise burst stimulus of approximately 500msec in length. The noise stimulus was white noise, low-pass filtered at 11 kHz.

The noise burst was presented from a small loudspeaker held by the experimenter who placed it at given spatial locations around the subject's head. The experimenter kept the source at a constant distance of approximately 18" from the center of the subject's head while varying the azimuth and elevation of the source to one of sixteen locations. Let [θ, ϕ] be azimuth and elevation coordinates, with [0°, 0°] defining straight ahead of the subject in the horizontal plane of the ears, and with positive azimuth angle proceeding to the right from the median plane and positive elevation angles proceeding upward from the horizontal plane. The sixteen possible source directions, all of which were in the subject's right hemifield, were formed by the fifteen combinations of θ = [0°, 45°, 90°, 135°, 180°] and ϕ = [–45°, 0°, 45°], in addition to the straight-up location at ϕ = 90°. These angles are easily reckoned by the experimenter without the aid of physical measurement scales. They are also familiar angles even to subjects who may not be accustomed to localization angle contemplation. Subjects were told of these 16 possible source locations and responded directly with azimuth and elevation angle estimates. After recording a subject's response on a trial, the experimenter then placed the hand-held source at the next location while avoiding any extraneous physical cues (acoustic, air motion).

A run consisted of two stimulus presentations from each of the sixteen locations, for 32 trials, with the constraint that all sixteen locations were tested in random order on the first sixteen trials and then again on the second set of sixteen. Two runs were typically conducted for each subject for each condition. All procedures were the same for all experimental conditions (helmet, microphone array, etc).

4.3.3.2 Subjective Qualitative Observations

A subjective quality assessment of selected COTS devices was performed including hear-through systems and earplugs. All hear-through devices were tested with the active hear-through system enabled. A total of three tests were performed in an acoustically controlled environment: 1) using live conversational speech, 2) using loudspeakers to reproduce sound sources and 3) a combination of (1) and (2).

In the first part of the test, live conversational speech was used in order to judge the behavior of COTS devices for speech communication. Judgments were based on the quality of speech, localization capabilities and overall sound quality.

The second part of the tests used three loudspeakers arranged in a triangle. Each loudspeaker reproduced a unique sound source of different frequency bands: a) 37 – 200 Hz, b) 4 – 12 kHz, c) 37 Hz – 20 kHz. Sound sources were reproduced at a level between 82 and 94 dB SPL (A-weighted), measured at the center of the triangular speaker arrangement. The evaluation focused on sound source localization, sound quality in three frequency ranges (low, mid and high) and spectral coloration.

Lastly, a combination of live speech communication and loudspeaker-reproduced sounds were presented to subjects wearing the selected COTS devices. Here, the evaluation focused on the behavior of the hear-through system with regards to speech intelligibility when other sound sources are present.

THIS PAGE IS INTENTIONALLY BLANK.

5 Results

This section describes the results in this study, including the numerical modeling exploration as well as prototype and COTS system evaluations, both quantitative and qualitative.

5.1 Modeling with Numerical Computation

The numerical computation was explored to approximate HRTFs for various head/helmet geometries. The goal was to proceed from an input geometry, set of microphone locations, and a beam direction and frequency, and to produce an estimate of the amplitude and phase of the sound pressure level variation at each microphone location. Repeating this calculation for various frequencies yields the estimated HRTF.

Only incident sound fields consisting of (single-frequency) plane waves were considered, corresponding to a point source at infinity in the given beam direction of arrival, as this case is sufficient for the evaluation of helmet geometries and microphone placements in terms of transparency of hearing. The surfaces of the helmet, head and relevant portions of the torso were assumed to be sound-hard. Under this hypothesis, the problem to be solved was to evaluate, at a fixed set of points (microphones), the solution of an exterior Neumann problem for the three-dimensional Helmholtz partial differential equations (PDE). Because this calculation must be repeated for many different frequencies and beam directions, a highly efficient algorithm was desired.

5.1.1 Background

Work began with a literature search for descriptions of previous approaches to acoustic scattering problems of this type. Most previous research in this area seems to follow the approach outlined above (including a fairly common use of the simple Neumann boundary condition), with several methods used to compute solutions to the Helmholtz equation. At the start of the project the expected most viable approach was some variation of the boundary-element method (BEM). The literature search generally confirmed this suspicion; while other techniques have been tried, they tend to be poorly suited to efficient implementation for exterior problems[9], e.g., finite-element methods (FEM), or difficult to apply to irregularly-shaped scattering bodies. One interesting alternative was discovered, the so-called infinite-element method, but this technique seemed to be less mature at present than the BEM, and potentially harder to apply to scatterers with complex geometries.

5.1.2 Method Selection

Because of time constraints prohibited extended evaluation of different approaches, BEM was pursued. Simple (so-called "direct") numerical solutions of exterior scattering problems via BEM are known to suffer from a physical lack of uniqueness at certain frequencies. The spacing between successive problematic frequencies tends to be smaller at higher frequencies, and as solutions valid in the relatively high range of frequencies that carry localization cues for human listeners were sought, a so-called indirect BEM scheme, which has no such lack of uniqueness, seemed advisable to use. Both direct and indirect BEM approaches express the solution to the scattering problem in terms of integrals of potentials taken over the surface of the scattering body, but the potentials used in indirect schemes contain terms arising from derivatives of the potentials used in direct schemes. So one pays a price for the uniqueness of the solution: the potentials to be integrated in an indirect scheme are hyper-singular (i.e., possess higher-order poles than those for direct schemes), a characteristic which complicates numerical evaluation of the surface integrals.

The BEM, like the FEM, replaces the PDE to be solved by a matrix equation for a vector of unknowns describing an approximate solution to the PDE. The solution process can thus be conceptually split into

[9] By "exterior problem", we mean convex surfaces which scatter sound.

two parts: first, coefficient matrices are generated, and then the resulting matrix equation must be solved. The first part requires that the surface of the scattering body be divided into panels, and for each pair of panels one or more integrals are computed over the product of the two panels. The second part can be handled by a standard algorithm such as GMRES (a conjugate-gradient procedure applicable to non-symmetric systems). Sophisticated variations of the BEM (based on wavelet decompositions, panel-clustering, or fast-multipole methods) exploit the asymptotic dependence of the surface potentials on distance to reduce the number or complexity of the integrals to be computed. However, they typically greatly increase the complexity of implementation of the integration phase, and sometimes also of the GMRES phase of the solution.

Early on, the fundamental decision was made to divide the scattering surface into a large number of simple panels, namely triangles. This division permits modeling of fairly arbitrary geometries relatively easily and greatly simplifies the surface integrals that must be computed. By comparison, to model any but the simplest helmet geometries using a small number of panels with tractable mathematical descriptions is difficult; moreover, use of large curved panels precludes analytic simplifications that can be applied to integrals over small flat ones. The disadvantages of this approach are that many more integrals must be computed and a larger resulting matrix equation solved, and that one must have some means of generating a model of the scattering surface as a collection of triangles.

The first problem can be addressed in either of two ways: by making computation of the individual panel integrals as efficient as possible, or by using one of the sophisticated BEM algorithms mentioned above. In fact both approaches are likely to be advisable. A simple BEM variant was attempted first, in order to produce a working software prototype as rapidly as possible. Then, time permitting, one of the more efficient variants would be implemented.

The second disadvantage is not severe. Geometric modelers capable of producing surface triangulations are commonly used for CAD/CAM and animation. Development of the BEM software began by obtaining a copy of the visualization toolkit (VTK), a free software package including such a modeler, and using it to produce high-resolution triangulations of spheres. Because the exact solution of the scattering problem for the sphere is known, it is useful in building a test case for the proposed software. VTK also permits construction and triangulation of other, more realistic geometries. In a final version of the HRTF software it may be desirable, for reasons of user convenience, to switch to another geometric modeler. As the modeler need not be coupled tightly to the remaining software components, it is unlikely that such a change would present any real problems.

The second phase of our effort thus concentrated on the production of software for the efficient numerical computation of the surface integrals arising in an indirect BEM using triangular panels, and on mathematical analysis of the relevant integrands with the goal of simplifying the inputs to the numerical integration as much as practically possible. The choice of flat triangular panels permits a great deal of mathematical simplification, but there is still need for carefully designed code. Initial software development was performed in Matlab™, using its built-in routine for multi-dimensional numerical integration. In search of greater efficiency, we re-implemented this code in C++ using similar algorithms; the resulting software was still too slow to be of practical use in a complete BEM package for our scattering problem.

5.1.3 Surface Integration Algorithms

Work began on development of faster surface integration algorithms in Matlab (but with an eye toward eventual implementation in C++). Simple Romberg and adaptive routines based on low-order cubature rules were written and tested. As their performance (in combined terms of speed and accuracy) did not seem sufficient for the intended application to the BEM, we then investigated the development of higher-order cubature rules, either for stand-alone use or for incorporation into an adaptive routine. This work remains unfinished.

5.2 *Acoustic-Measurement-Based Error Metrics*

Measurements of the various systems were divided between the West and East Coast laboratories for efficiencies. Both laboratories used identical AuSIM HeadZap software for measurements, but the physical measurement configuration varied. The reference systems were also different: KEMAR vs. Bruel & Kjaer's HATS. Thus we have divided the prototype approach results into two groups, each compared back to the respective references.

All commercial systems, with the exception of the Peltor COMTAC device, were measured against HATS and presented as a group separate from the prototype results.

5.2.1 Commercial Head-Borne Systems

5.2.1.1 Active Hear-Through Hearing Protection Systems

Figure 44 displays the error metrics calculated on the measured data of hear-through devices with activated gain control; it also displays the error metrics for helmets with and without accessories. All measurements used the HATS as the reference. The data includes ITD, ILD and magnitude error measurements.

The ITD errors on the hear-through systems were varied. The ITDs were highly dependent on the placement of the microphones on the muffs. Larger muffs with microphones placed on the outer edge resulted in larger ITDs and, thus, a greater error. The lowest error metric was seen on the Bilsom and Leighting systems. The Sordin system experienced the most significant error rate. This large error could have been partially caused by a phase discrepancy between the left and right channels.

There is a substantial variation in the ILD errors across all hear-through systems. The high level of error is most likely due to the independent AGC on the left and right channels. The AGC condition should be completely isolated in future testing to evaluate ILD without gain control. The poorest performance is seen in the Remington 2000 while the Leighting system had the lowest error rate.

It is not surprising to see the high level of magnitude error caused by the hear-through systems. The independently measured spectral characteristics of the hear-through systems show substantial coloration, particularly at high frequencies. This coloration is obvious in the measured HRTFs. Most systems show similar errors, except for the Remington 2000, which showed a substantially higher error rate.

5.2.1.2 Helmets and Accessories

Measurements of the MICH helmet show relatively low errors in ε_{ITD}, ε_{ILD}, and ε_{mag} when the helmet was measured with no accessories. Even lowering the night-vision goggles (NVG) had little impact on the result. Measurements of the MICH helmet in combination with the chem-bio mask reveal a substantial increase in error in the ITDs with no effect on the frequency response magnitude. The MICH helmet combined with goggles and muffs shows the poorest result, with ITD and ILD errors almost doubled. The Scorpion R2 helmet was measured in three different settings: (1) all accessories and muffs attached, (2) all accessories and muffs removed, and (3) muffs removed but accessories attached. As can be seen in Figure 44, the Scorpion helmet results in a relatively small error rate with the muffs removed. Helmet accessories have little effect on the measurements. However, muffs have the greatest effect on ε_{ITD}, ε_{ILD}, and ε_{mag}, more than doubling all three values.

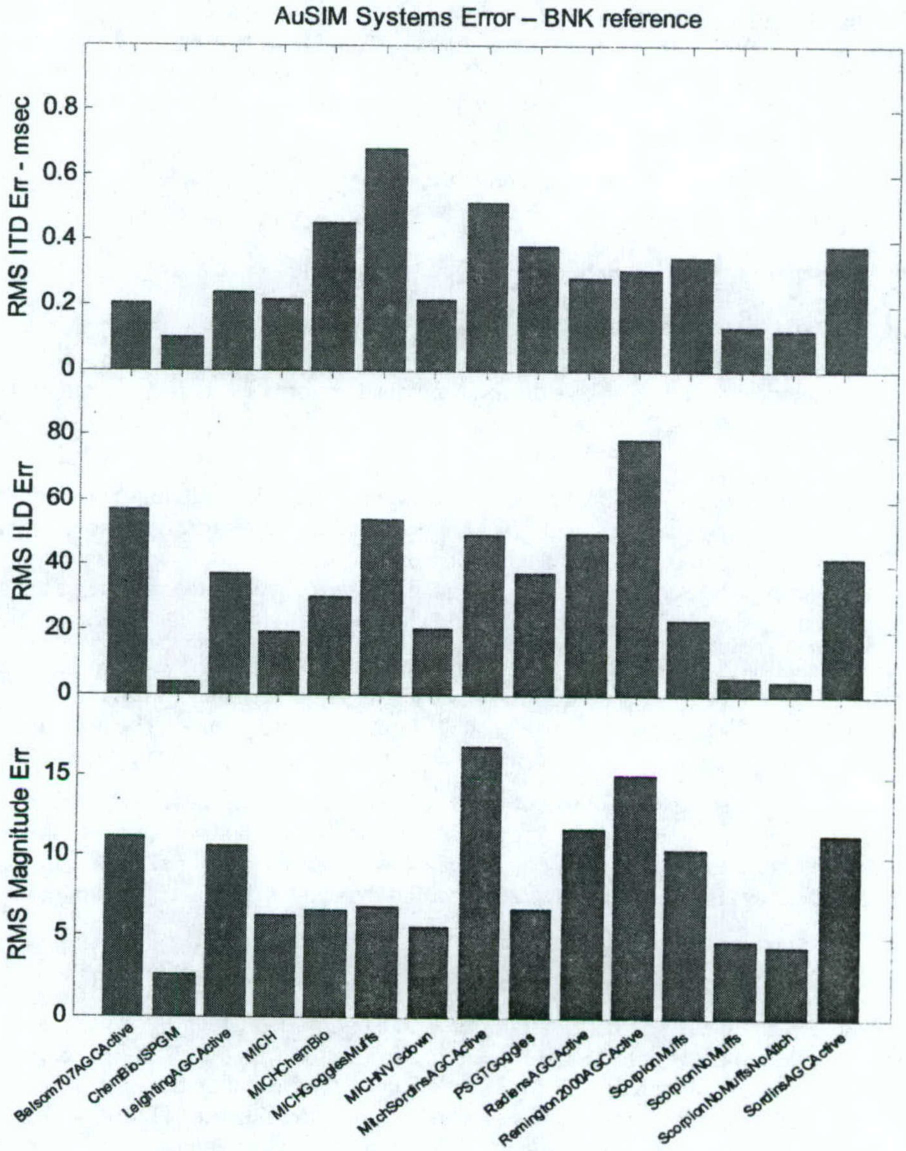


Figure 44: Error metric results for measurements using B&K HATS as reference,

5.2.2 East Coast Laboratory Prototypes

5.2.2.1 Hidden Concha, Simulated Pinnae and Microphone Array

Figure 45 displays the three error metrics ε_{ITD}, ε_{ILD}, and ε_{mag} calculated for the hidden concha (HC), simulated-pinnae (SP), and general microphone array (GA) test systems. The abbreviations for the various reference and test systems are identified in Table 5. For the Reference and HC systems, these metrics were calculated using KEMAR as the standard. For the SP and GA systems, the error metrics compared the responses through four system implementations to their respective custom targets (the three subjects and KEMAR). The error bars on results for those systems are the standard deviations across the four custom implementations.

RMS ITD Error

The RMS ITD error, ε_{ITD}, is essentially the same for the SP and GA systems. This result is predictable for the SP systems given the fact that the microphones providing signals to one ear were located on the muff at that ear. Given that the distance between muffs is slightly larger than between ears, and that it was approximately the same for those systems, the same error would be expected. Similarly, as described in Section 3.2.2.5, the GA systems divide the processing into low-pass and high-pass components. The low-pass component is based upon a single-microphone SP system, and so the GA systems exhibit similar ITD characteristics to the SP ones. The HC system showed the lowest ITD error of all systems.

RMS ILD Error

The ILD error, ε_{ILD} (which is proportional to a dB scale), is also roughly constant across experimental systems. The large error for the PEL headset is a result of the independent AGC in the two muffs. During the DTF-measurement process, sources in some directions would trigger the gain control at one side but not at the other, leading to large interaural level differences. Again, the HC system had the lowest error metric of all systems.

RMS Magnitude Error

The RMS magnitude error, ε_{mag} (which is proportional to a dB^2 scale), is intended to capture the features in the direction-dependent spectral magnitude response. For this metric, the SP systems all had roughly the same error, while the GA systems' error was about twice as large. This result reflects the inferior LSE design metric used for the GA systems as opposed to the LSM metric used for the SP systems. The PEL headset also had a very large error, which is again partially due to the AGC. The fact that the HC system had relatively small RMS magnitude error is expected since the system's conchae were molded from KEMAR's conchae.

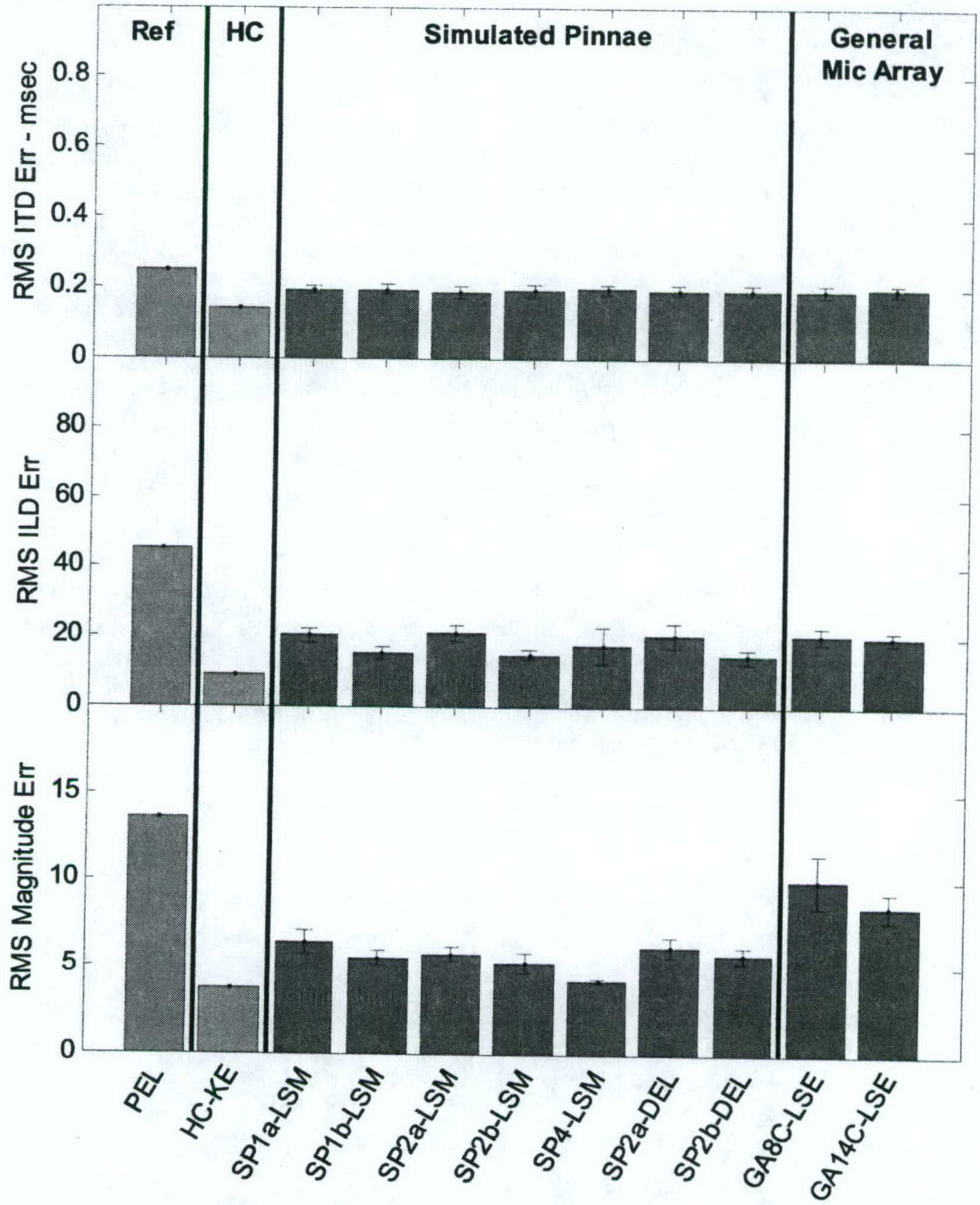


Figure 45: Error metric results for hidden concha (HC), simulated-pinnae (SP), and general microphone array (GA) systems.

5.2.3 West Coast Laboratory Prototype Systems

The results of the error metrics of prototype systems are shown in

Figure 46. The measurements were taken and analyzed using the subject ARO as reference.

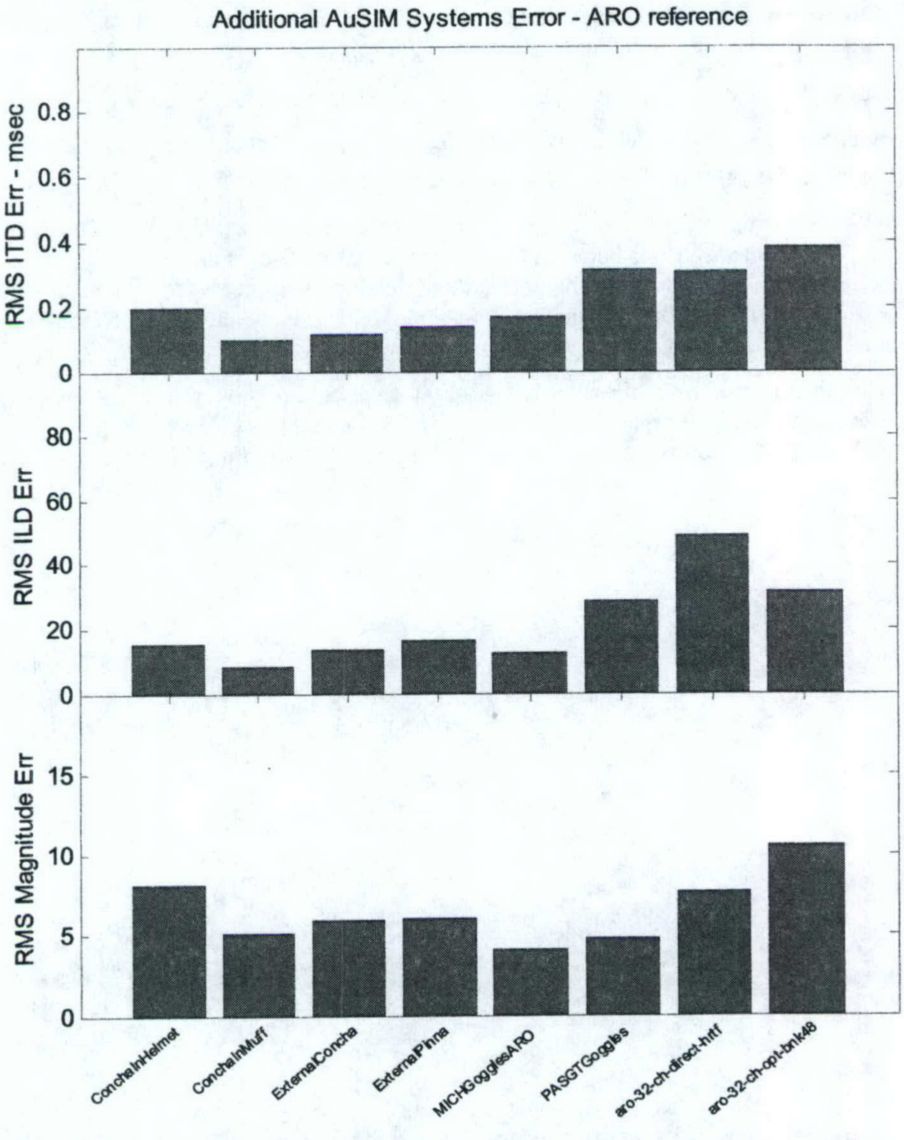


Figure 46: Error metric results for measurements using ARO as reference. Please note that for this graph: ConchaInHelmet = Concha-In-R2Helmet, ConchaInMuff = Concha-In-R2Muff, ExternalConcha = HardPinnaMuff, and ExternalPinna = SoftPinnaMuff.

The ITD errors, ε_{ITD}, were not as expected. The system with the concha placed in the crown of the Scorpion R2 helmet ("Concha-In-R2Helmet") was especially designed to replicate the interaural distance and thus maintain normal ITD cues. However the system with the concha in the muff of the Scorpion R2 helmet ("Concha-In-R2Muff") resulted in the smallest ε_{ITD}. The time difference is not based on the interaural distance alone, but the total length to circumvent the shadowing. Because the Concha-In-R2Helmet has significantly different shadowing effects for non-zero elevations, the ε_{ITD} may have been skewed. The soft and hard pinna systems had approximately the same distance offset from the ears and resulted in similar ε_{ITD} that were somewhat larger than the Concha-In-R2Muff case. The system containing the concha placed on either side of the R2 helmet resulted in a much greater distance between the left and right channels and thus increased the ITD error.

In binaural systems, the ε_{ITD} and ε_{ILD} tend to be somewhat related, as the ILD was dependent on the head shadow. This relationship between the ε_{ITD} and ε_{ILD} is evident in results shown in Figure 46. Again, the concha in muff system had the lowest ILD error.

The RMS magnitude errors were similar for the Concha-In-R2Muff, HardPinnaMuff, and SoftPinnaMuff systems. The small variations between the differences can be attributed to the dissimilarities of the external ear shape. The common characteristic of the three systems is the relationship between the system and the shoulders. When the concha was placed in the helmet the shoulder reflections were much different. The alteration of this important cue resulted in higher error metrics.

For the 32-channel array, the direct HRTF filtering gave better ITD error than the optimized filter because it had a more stable phase response. The direct HRTF had a poor ILD error because the microphones contained overlap in their directivity patterns. The use of the optimized filter was an attempt to minimize this error.

5.3 Localization Test Performance

The results of the behavioral measurements of sound localization for the various reference and experimental Transparent Hearing Systems are presented in Figure 47. Each bar and associated error bar are the average and standard deviation, respectively, of the error measures taken on two repetitions of the 32-item test for three subjects. Each error measure, in turn, is the RMS average error over source locations. One subject was unable to complete testing and, as a consequence, the results for the OE, OE-H, PEL, SP1a-LSM, and SP1b-LSM systems include data from only one trial with this subject, and the results for the GA8C-LSE and GA14C-LSE systems include no data from this subject. The SP and GA systems were designed and tested for two different desired HRTFs for each subject: KEMAR HRTFs and the subjects' own, custom-measured HRTFs. These 'KEMAR' and 'custom' conditions are distinguished by blue and red bars.

5.3.1 Error Measures Employed

Three different error measures are presented in the three panels of Figure 47. The error measure used in the upper panel of Figure 47 is the angle error between the ideal source location and the subject's response coordinates. This measure includes errors in both azimuth and elevation and does not include correction for front/back confusion. For a listener responding randomly with one of the allowed locations, the expected value of this error measure is approximately 87°.

The error measure plotted in the middle panel of Figure 47 is the error in azimuthal angle only. If, for example, a source was presented at an azimuth of 90° and an elevation of -45°, and the subject responded "90° azimuth, +45° elevation", there would be zero error. A response of "straight up" was regarded as correct for any azimuth. The expected value of this error measure for a random-response is 84°.

The error measure used in the lower panel of Figure 47 is the error in elevation. Analogous to the previous measure, an elevation error is measured only by the deviation in the elevation dimension, regardless of the azimuth component. The expected value of this error measure for a random response is 51°.

Despite the differences in the error measures, the trends are remarkably similar for all three. The results can therefore be discussed in common.

5.3.2 Localization Performance

The best performance over systems and conditions was obtained with OE and OE-H, which appeared to be equivalent. The worst performance over all systems was with the PEL system. With that system, listeners achieved no better than chance performance on elevation perception, and only slightly better than chance on azimuth. This is not to say, of course, that there was no structure in their error patterns. But there were too few responses per cell per listener, and the responses were too idiosyncratic for each listener to construct confusion matrices.

Among the experimental Transparent Hearing Systems, the single-microphone systems, SP1a-LSM and SP1b-LSM, also gave poor elevation performance; but their azimuth performance was clearly better than that with the PEL. In fact, all of the SP systems gave roughly equivalent azimuth errors; these were outperformed in azimuth error as a class by the two GA systems.

The overall best experimental systems were the two using two microphones on each muff with a simple delay-and-sum algorithm (SP2a-DEL and SP2b-DEL). It can be seen that this superiority results primarily from their better performance in elevation. The GA systems produced large elevation errors, comparable to those of the PEL.

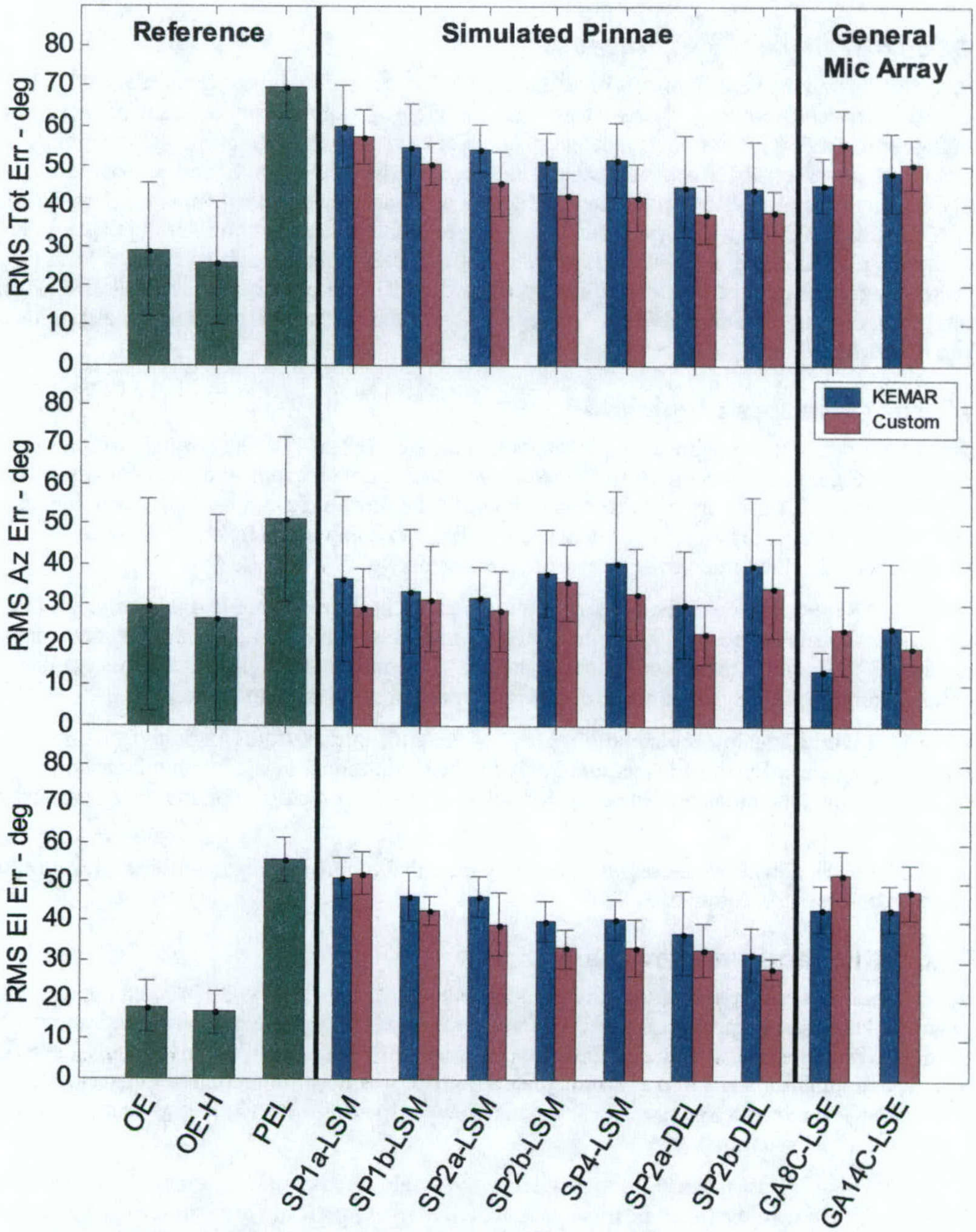


Figure 47: Subject localization test error results showing RMS total average error, RMS azimuth location error, and RMS elevation error.

5.3.3 Front/Back Reversals

A large and frequent type of sound localization error evident in both the open-ear reference systems and in the various test systems is a front/back reversal. If a source has azimuth θ, there are frequent erroneous responses at 180-θ. These errors result because the interaural cues are similar for sources at front-back-symmetric positions[10].

Figure 48 shows the average percentage of stimulus presentations that resulted in front-back confusions, with error bars indicating the maximum and minimum percentage over all test subjects and presentations. The PEL system exhibited the highest level of front-back confusions, while the GA systems exhibited the lowest. In general, the SP and GA systems resulted in confusion levels similar to the OE and OE-H reference systems.

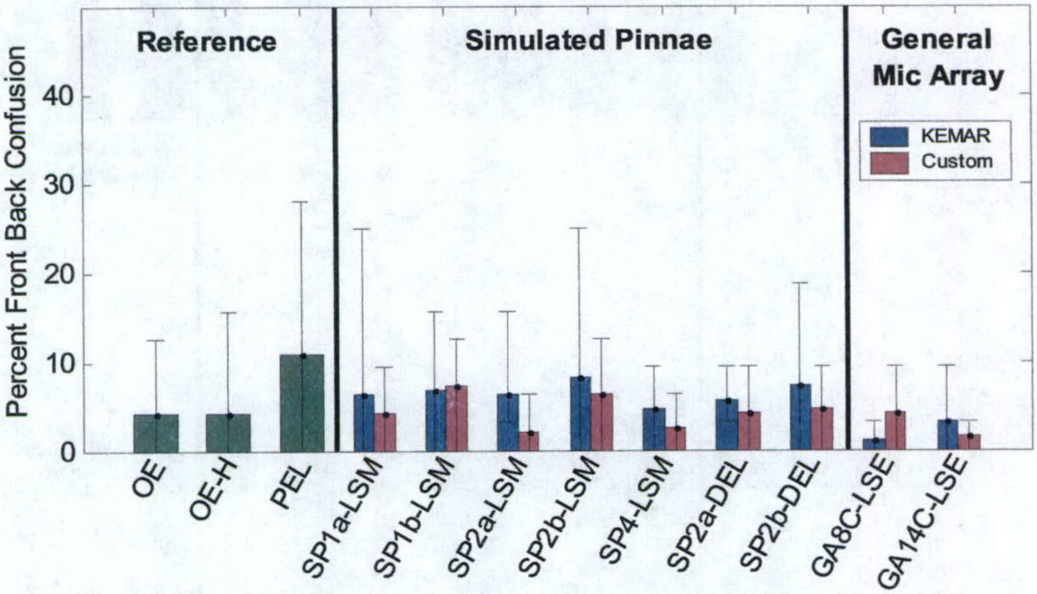


Figure 48: Average percentage of trials that resulted in front-back confusions. Error bars indicate maximum and minimum confusion percentage for each system.

The results of Figure 47 have been re-plotted in Figure 49 with front-back confusions removed. Because the removal of front-back confusions has no effect on elevation errors, the data in the lower panel of Figure 49 are the same as in Figure 47. The expected random-response performance in azimuth error is now 50° and in total error is 67°.

The overall profile of performance across systems remains unchanged. All systems perform more poorly than either OE or OE-H on azimuthal localization. The systems providing the best elevation performance are the wideband delay-and-sum algorithms, SP2a-DEL and SP2b-DEL.

[10] Front/back reversals are a common phenomenon relating to the discussion on the Cones of Confusion in section 2.1.2.1.

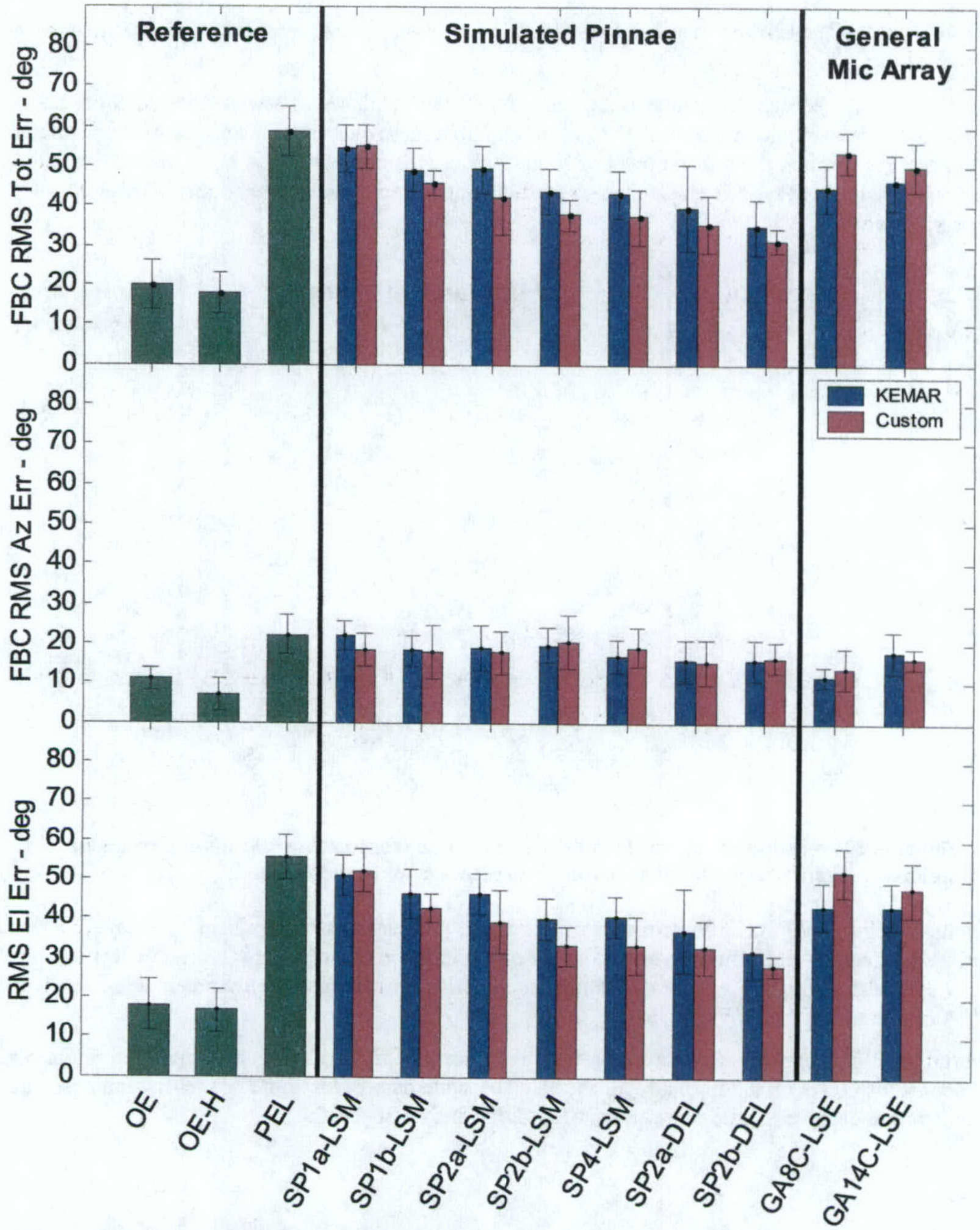


Figure 49: Error results for the subject localization test, with front-back reversal errors corrected, showing RMS total average error, RMS azimuth location error, and RMS elevation error.

5.3.4 KEMAR versus Custom

Comparisons of systems designed to match KEMAR's DTFs to those designed to match the individual user's DTFs showed a small but consistent advantage to the custom designs for localization in both elevation and azimuth. The advantage is most pronounced in the SP approaches with front/back confusions included, as shown in Figure 47. With the exception of the SP1b-LSM and the G8C-LSE systems, subjects experienced fewer confusions with the custom-HRTF SP and GA systems than with the KEMAR HRTF systems.

5.3.5 Analysis

One would like to see, of course, a clear correlation between the physical error metrics from the previous section and the results of the localization tests in this section. With respect to interaural cues, azimuth localization error in Figure 47 (or Figure 49) appears roughly constant across experimental systems, consistent with the constant error metrics for ITD, the cue that is most important for azimuth localization. With respect to cues for elevation, however, the relations are not at all clear. For example, there is no indication in the approximately constant RMS Magnitude error metrics for SP systems that correlates with the trends seen in the localization results. The large error metrics for the GA systems have no counterpart in the localization data. There is, however, consistency between the large RMS Magnitude error metric and poor elevation localization for the PEL headset.

5.4 Subjective COTS Quality Assessment

Selected COTS systems were tested, as described in Section 4.3.3.2, for sound quality, overall system performance, spatial cues, and comfort. Results of the subjective assessment tests show that the best overall system with regards to comfort, sound quality, spectral response and localization cues is the Howard Leight Pro-Ears Leightning system. The response of the low frequencies was considered to be attenuated but the response of the mid and high frequencies was good.

Conversely, the Sordin Supreme III hear-through system was judged to be unacceptable in the current configuration. Although the spectral response of the system was good, the spatial cues and image coherence was very poor. Further physical testing revealed a phase reversal in one of the channels and thus caused severe sound image distortions.

When tested in the presence of other sound sources, speech intelligibility was very poor for all systems. However, the Radians Pro-Amp Electronic Earmuffs were perceived to have the best speech intelligibility.

Tested earplugs included foam plugs, Silencio plugs, and the brown/yellow Combat Arms Earplugs[11]. The yellow end of the Combat Arms Earplugs was judged to be the best for speech intelligibility with good mid and high frequency spectral response. The brown end of the same plugs was perceived to have a substantially high frequency attenuation thus leading to a dampened sound. The Silencio plugs were considered to be as good for speech as the yellow end of the Combat Arms Earplugs but were less comfortable.

Results of the tests performed are summarized in the table given in Appendix E.

[11] Recall from section 3.1.2 that the brown end is a total plug, while the yellow end is a passive hear-through protector.

THIS PAGE IS INTENTIONALLY BLANK.

6 Headgear Design Guide

To guide the design of future headgear, the design guide should explore issues relevant to the human factors, present alternative ideas, and reflect the analysis of the data collected. However, the analysis of the data was beyond the scope of the present project, limiting the concrete guidance derivable from the extensive testing done in this project. Still, some critical issues relevant to ear-protecting headgear, such as heat-dissipation, were explored and some ideas are presented here.

The images on the following pages are sketches from the design exploration with annotation.

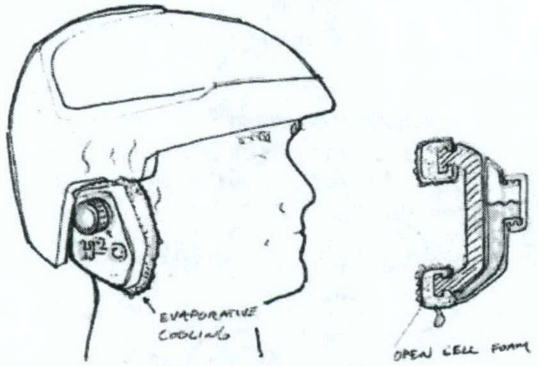


Figure 50: **Cool_1** Liquid Reservoir and wicking arrangement provide temperature differential via evaporative cooling

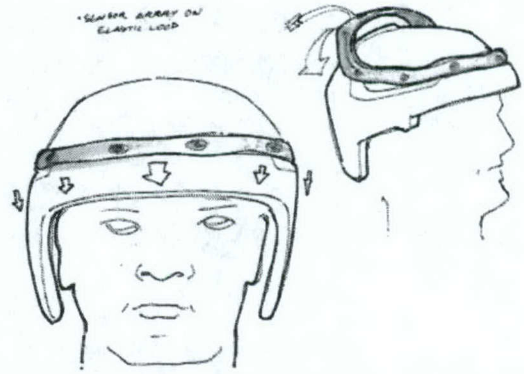


Figure 51: **MicPlacement_1** Radial mic array on retrofit conforming harness

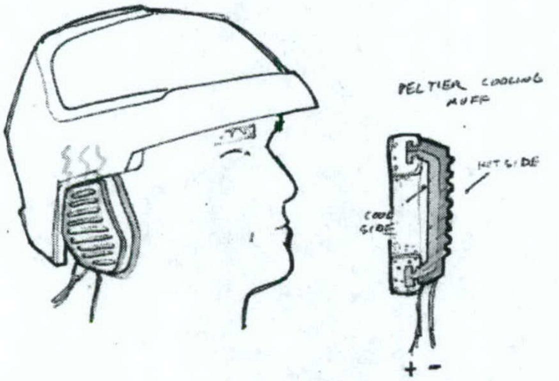


Figure 52: **Cool_2** Active cooling systems can be employed – including this design for a Peltier solid state heat pump

Figure 53: **MicPlacement_2** Mics located on neck strap to align with inter-aural distance and provide ground sensitivity

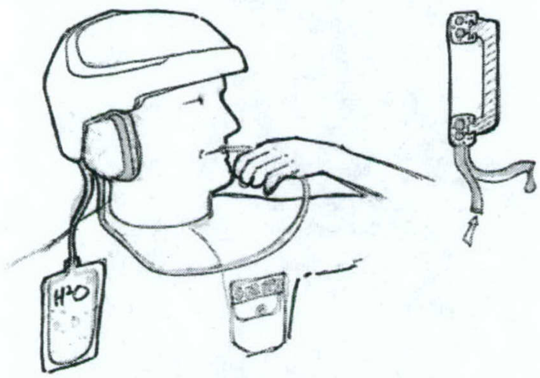


Figure 54: **Cool_3** Hydration water used for muff cooling

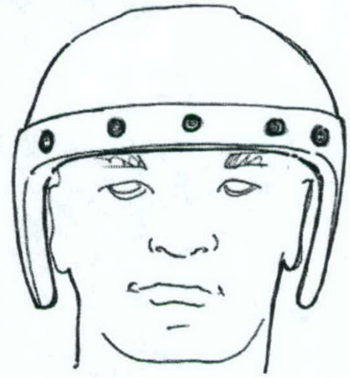


Figure 55: **MicPlacement_3** Radial mic array naturally located along existing R2 channel

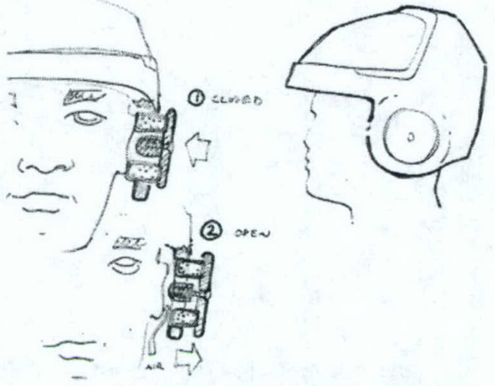


Figure 56: **OpenVent_1** Open air, non-sealing design can be slapped shut to provide an acoustic seal

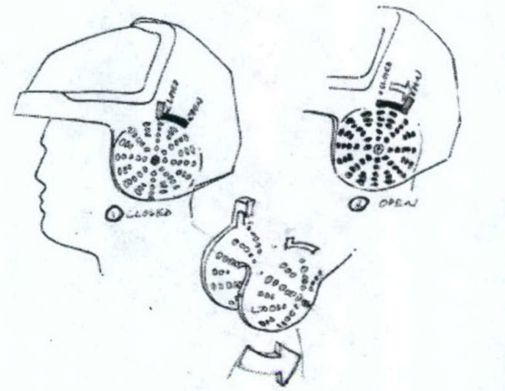


Figure 57: **OpenVent_4** Radial pattern of open air vents can be quickly closed via small angular twist to provide cooling or muting choices with natural hearing when open

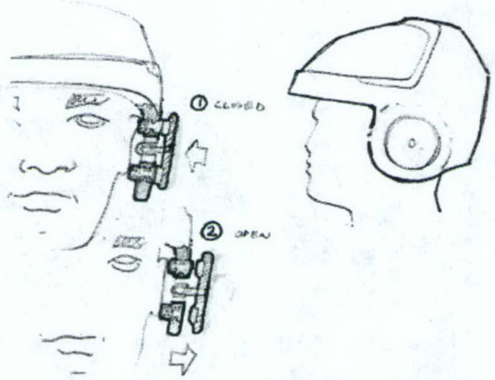


Figure 58: **OpenVent_2** Slap shut mechanism allows for rapid selection of cooling or muting

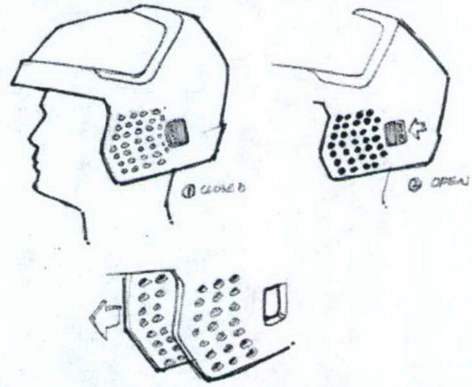


Figure 59: **OpenVent_5** Linear array of vents can be slid shut to provide hearing protection, or opened for cooling and natural hearing

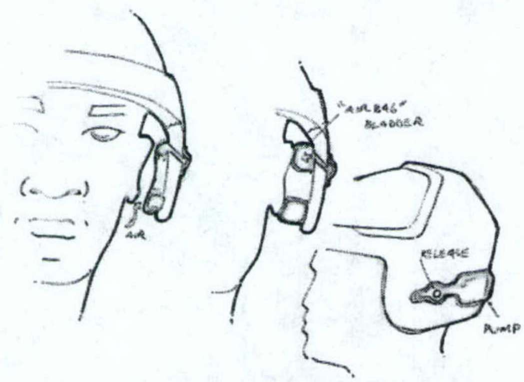


Figure 60: **OpenVent_3** Acoustic muting via hand pumped, or micro-airbag charge

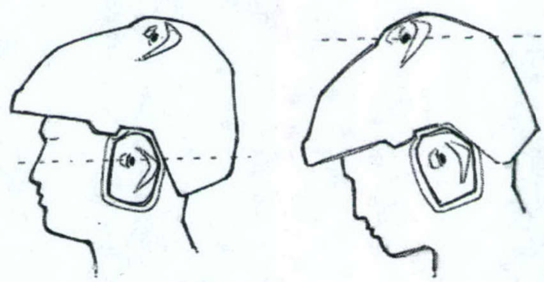


Figure 61: **DualSensor_1** Secondary Pinnae angled to actuate upon 'attentive' head pose

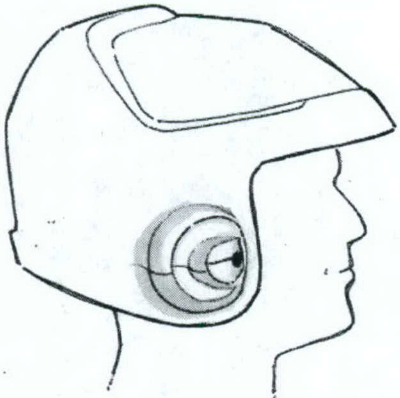


Figure 62: **StylizedPinnae_1** Mechanical pinnae forms styled into the helmet features.

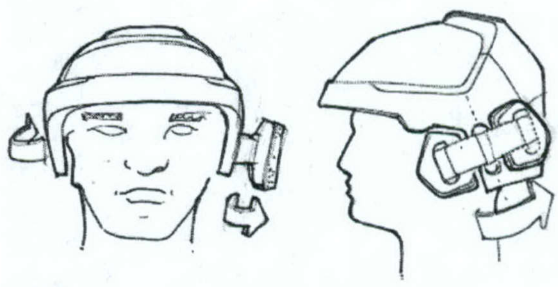


Figure 63: **SelfStowing_1** Out of the way, Muff Retention

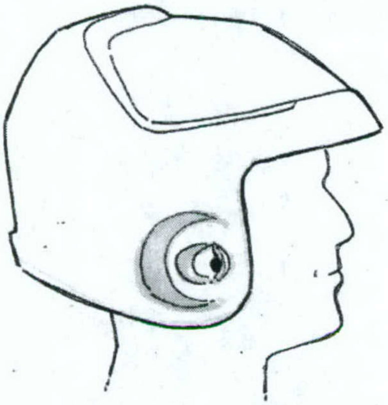


Figure 64: **StylizedPinnae_2** Alternative mechanical pinnae forms styled into the helmet features

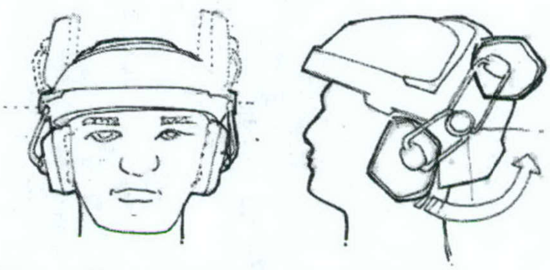


Figure 65: **SelfStowing_2** Muff retention option

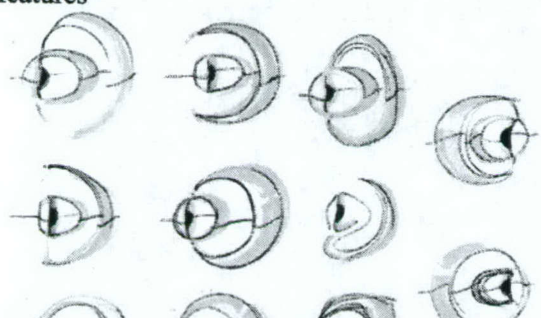


Figure 66: **StylizedPinnae_3** Family of stylized pinnae choices

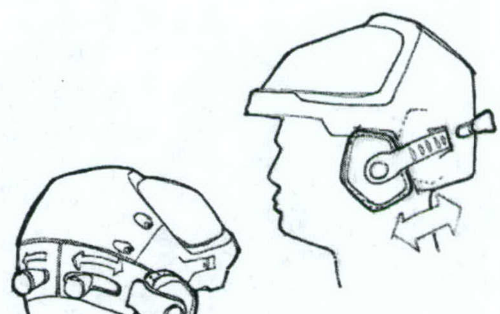


Figure 67: **SelfStowing_3** Muffs retained via retractable flexible strip

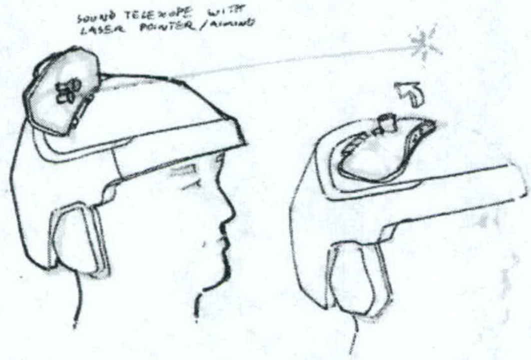


Figure 68: **SuperFocus_1** Physically based super-hearing option

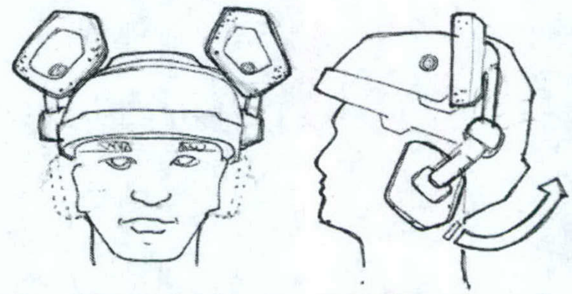


Figure 69: **SuperStowing_2** Muffs swing up to collect sound

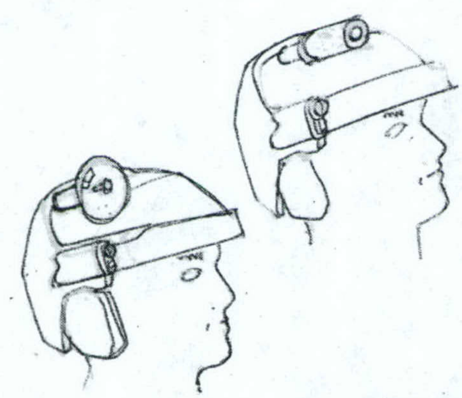


Figure 70: **SuperFocus_2** Super hearing options with metaphorical forms

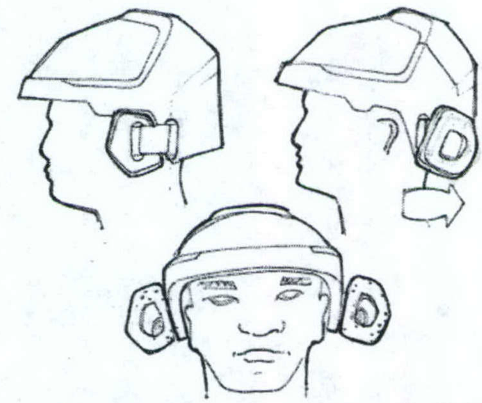


Figure 71: **SuperStowing_3** Tri-mode design: muffs on; swung back and out of the way; open to act as super-hearing acoustic collectors

Figure 72: **SuperStowing_1** Muffs act as super-hearing collectors when opened

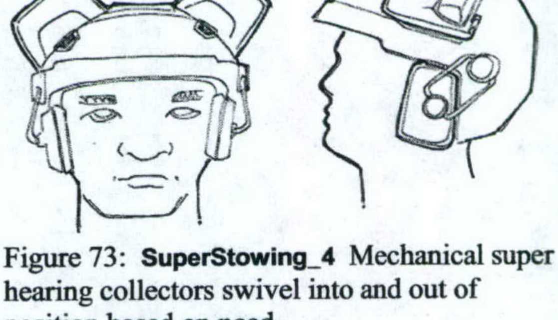


Figure 73: **SuperStowing_4** Mechanical super hearing collectors swivel into and out of position based on need

THIS PAGE IS INTENTIONALLY BLANK.

7 Final Remarks

7.1 Discussion

This study was undertaken to explore technology for providing transparent hearing to the soldier who needs to "hear through" their hearing protectors. Of primary concern in this study was the ability of a Transparent Hearing System to provide accurate sound localization performance. Several commercial hear-through hearing protectors were obtained and evaluated, and several experimental approaches were explored. Systems were evaluated in terms of numerical error metrics and in terms of a basic localization test. We summarize and discuss the results of all these aspects of the study in the following sections.

7.1.1 Simulated Pinnae Systems

Several variants of these systems were implemented and tested. These employed one, two, or four microphones placed on each muff, contributing signals for that ear only. The microphone location aspect of the design guaranteed a good approximation to the natural interaural cues, ITD and ILD. As a result, azimuth errors for those systems were within a factor of two of open-ear error rates. There was no evident difference in azimuth errors between system variants of purely microphone position (SP1a-LSM and SP1b-LSM).

The one-microphone variants of simulated pinnae systems gave relatively poor elevation perception, as expected. Elevation performance is best with the two-microphone systems that use delay-and-sum processing (SP2a-DEL and SP2b-DEL), where the delay was chosen to match the desired DTFs. Two- and four-microphone variants of simulated pinnae systems that were designed based on algorithmic search for best filter parameters (SP2a-LSM, SP2b-LSM, and SP4-LSM) produced larger elevation errors. This result highlights the difficulty of defining an error metric and a search procedure that effectively captures and optimizes the important features in the system response. Since the 2-microphone LSM variants use the same microphones as their counterpart DEL variants (and the SP4-LSM uses all four of them), an effective metric and search algorithm that was closely related to the important factors for localization should have produced behavioral error rates for the LSM variants equal to or smaller than for the DEL variants.

7.1.2 General Array Systems

These systems used 8 or 14 microphones with filtering optimized to match target DTF responses according to the design metrics. In lateral localization, these systems performed as well as, or better than, the simpler simulated pinnae systems[12]. For localization in the elevation dimension, however, these systems' error rates were among the largest. Again, we point to the difficulty in performing the automatic search in designing these systems as the primary problem.

7.1.3 32-Channel Apparatus

The 32-channel microphone array system was the most complex Transparent Hearing prototype developed. Its multi-microphone array setup lends itself very well to future expansions of supernormal capabilities.

Although the performance of this system was very promising, there were several potential problems that limited the performance of this approach. First of all, this approach relied on having some independence between the microphones, so ideally each of the microphones would have a fairly tight directional pattern so that there would be minimal overlap between the responses of neighboring microphones. Second,

---

[12] Recall in section 4.2.4.6 that the general microphone array systems used binaural omni-directional microphones in the low-frequency region.

there was a large gap in microphone coverage around the listener's face. This gap caused sounds that originate in this region to be perceived as coming from directions other then directly ahead. Third, there was an additional exaggeration of the time delay of the sound, since the ITD of the HRTF was concatenated with the inherent delay due to the microphone location. Due to the overlap in microphone responses, there was a large amplification of sound reflections in the room. This amplification made everything sound a little more "live" and sometimes made it hard to hear the direct path to the sound over the reflections of the sound. Despite these limitations, this system sounded unexpectedly good.

Using DTF-to-HRTF filter optimization for the 32-channel system gave reasonable results. However, the results were limited due to the difficulty in matching both magnitude and phase in the complex plane. Therefore, the ITD's did not match the desired delays very well, and often the magnitude of the contralateral ear did not match well either. Consequently, the Direct HRTF system sounds perceptually better, even though its measured response was not as "close" to the ideal HRTF response.

Using Beam optimization for the 32-channel system, results were disappointing. Given the good numerical directional response of the beams, one would expect this approach to eliminate or minimize most limitations of the 32-channel system. Further investigation is warranted.

7.1.4 Sound-field Microphone Apparatus

The minimum phase optimization gave very promising results. The captured sound-field sounded very natural and had very good informal localization. This prototype was completed too late to be included in the acoustic evaluation comparisons for this study. An exaggeration of ITD cues was expected since the microphones are located on the outside of the muff.

7.1.5 Physical Pinnae Systems

The human-replica physical pinna system performed subjectively well, but was not aesthetically pleasing. Even with the wind-screen covers, the system was proportionally too large. The alternative physical pinna systems produced directional cues, but did not yield immediate externalization. Externalization may be gained by training or adaptation, which would allow the user to adopt the foreign cues. Further study is warranted.

For the goal of finding a quicker means of exploring alternative pinna shapes, the work in mathematically modeling the pinna proved to be beyond the scope of current methods and tools. A physical pinna solution may exist, but more work must be done to focus on a set of solutions.

7.1.6 Commercial Systems

The Peltor COM-TAC system evaluated by Sensimetrics was uniformly the worst system tested in terms of both error metrics and behavioral results. At least a partial source of the problems with that headset was the interaurally-independent AGC.

The Sordin Supreme III displayed particularly poor sound quality. The Sordin and Peltor systems were very similar in design and performance. Although the frequency response appeared visually good, the system sounded unnatural. When tested in isolation, the Sordins presented a phase reversal between the left and right channels.

The Leightning AGC system was perceived to be the best overall system. The system offered a good spectral response and good spatial cues, and the earmuffs were comfortable to wear. The Remington system sound quality was perceived to be equally good, however with a slower AGC response.

The Bilsom system was perceived as having a generally good frequency response, but was weak in the low frequency range. A strong directionally dependent coloration was perceived and a distortion was heard in the AGC.

The Radians proved to be a good system, particularly for speech. Generally, a poor low frequency response and a boost in the 10 kHz range made the audio sound tinny. There was some distortion present with louder sources at close distances.

The COTS devices tested were a sparse sampling of the available systems. A more robust study would test more available devices. Active in-ear devices were omitted from the study. The AGC component creates a challenge in testing so as to isolate problems at AGC-activated levels versus normal hear-through sound levels.

7.1.7 Other Considerations

The majority of the present work centered on the pursuit of the approaches reported in the previous sections. However, stated project goals included gathering knowledge of various issues other than acoustic performance. This section discusses what was learned about some of these issues.

7.1.7.1 Cost

The current study did not advance any of the prototypes far enough to make useful cost estimates. The existing active commercial systems cost between $60 and $2500, with most quality hearing protectors above $300. The most costly devices are active in-ear, custom-molded hearing/communication plugs. The prototype devices in this study have significant variance in complexity. The most costly variable is the number of microphones, which scales analog support circuitry and digital processing. If the winning device can have a derivative commercial product (non-military), then high-integration can greatly reduce per unit costs for large quantities.

7.1.7.2 Compatibility

A factor that should be kept in mind in assessing the options for Transparent Hearing Systems is the potential offered for added functionality beyond the immediate task of hearing transparently through head-gear.

Accessory Headgear

Acoustic data was collected on various head-gear accessories and head-gear combinations. Some data showed significant disruption to spatial cues. Due to time constraints, this study was not able to perform enough analysis on the data to derive specific compatibility guidelines.

Advanced Auditory Displays

Results in the localization studies showed improved performance with custom versus generic HRTF's. This observation points to the importance of potentially matching the hear-through system to a user's open ear cues. A complete audio system integration, which includes other auditory displays such as voice communications and alerts, may also need to study cue-matching on a user-specific basis. If not individualized, an auditory display may need to match the cues of the Transparent Hearing System.

Advanced Augmented Hearing
Microphone array solutions may be used to provide enhanced directional hearing to the user: the signals from the microphones mounted on the muff or the helmet can be combined to form a directional filter that is more sensitive in a desired "look" direction than in other directions. Figure 74 shows directivity indices computed for the various array configurations used in the simulated-pinnae and general microphone array systems.

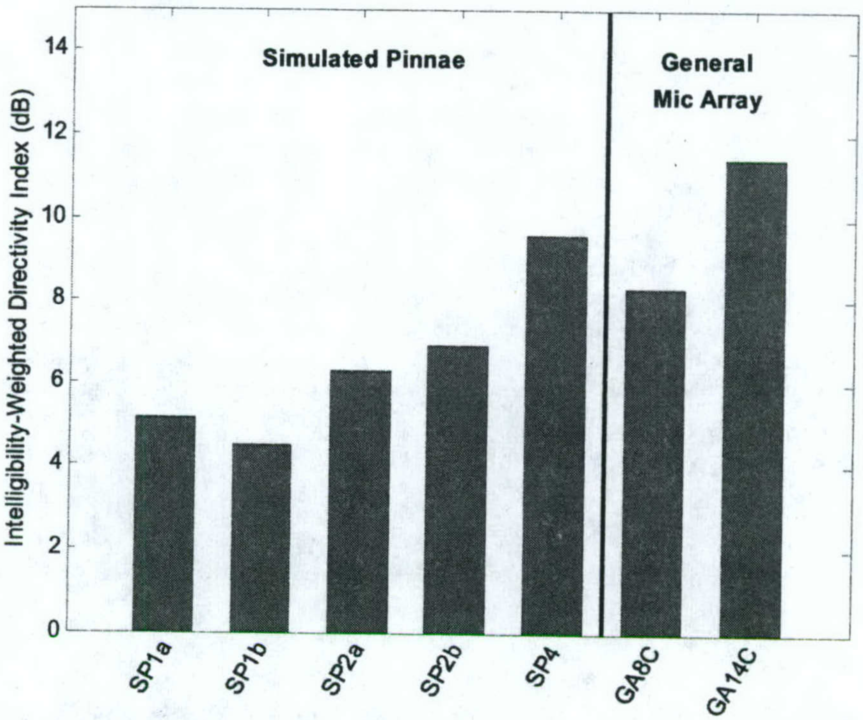


Figure 74: Intelligibility-weighted directivity indices for microphone configurations used in the simulated-pinna and general microphone array test.

7.1.7.3 Performance Specifications

Latency
When considering a transparent hearing system, total system latency refers to the time elapsed between the incidence of the sound arrival at the system and the sound delivery to the listener's biological hearing system. Any digital processing system will inherently contain latency due to some minimal requirement of processing time. The human auditory system is incapable of perceiving latencies below a threshold, often resulting in acoustic event fusion [19]. However, the effect of increasing system latency may result in inaccurate localization performance [139], unnatural perception of the acoustic environment, and a degraded sense of interactivity with the environment. In research pertaining to virtual environments, there is a widespread belief that this threshold for latency is not below 15msec [27]. Therefore a vital metric for a transparent hearing system is the maximum latency before the listener's performance becomes affected or the listener perceives latency. Due to time constraints, this study deferred the determination of this important metric.

Controlled Path Attenuation
The Transparent Hearing System relies on the elimination of the direct path so that it does not interfere with the processed signal. True and complete elimination of the direct path may not be possible or

needed for the purpose of Transparent Hearing. Rather, what should be considered is the perceptual elimination of the direct path. Thus, a vital guideline for transparent hearing system design should be how much acoustic attenuation is required to "control" the direct/uncontrolled path to the point of psychoacoustic elimination. Due to time constraints, this study deferred the determination of this important specification.

7.1.7.4 Plugs vs. Muffs

When designing a transparent hearing system, a major consideration is the physical configuration used for the transmission of the acoustic signal and attenuation of noise. Earplugs and earmuffs are the most common configurations for this task. Both configurations have their advantages and shortcomings in areas including hearing protection, comfort, sound quality, and hygiene. This study focused primarily on circumaural, sealed earmuffs. An optimal solution may depend on task and use definition.

7.1.7.5 Task definition for Evaluation

By project definition, the deciding factor of a successful Transparent Hearing System is the ability of users to perform a critical task equally well with a Transparent Hearing System as under the open-ear condition. An operational task that exemplifies current headgear-related hearing problems should be considered. This study did not identify any task or collection of tasks from the user community.

The resolution of such a task or set of tasks requires significantly more investigation than this study anticipated. A strong collaboration from the user community is required.

7.1.7.6 Near-field vs. Far-field Evaluation

Physical acoustics defines the near-field to be the region of space within a fraction of a wavelength away from a sound source, thus varying greatly with frequency. In terms of human localization, the near-field is accepted to be an area of space within 1 meter from the center of a listener's head, and the far-field is the space that is more than 1 meter away from the listener. From a localization point of view, the near-field is important as it is the only space where localization cues change as a function of distance. In the near-field, the head-shadowing effect is exaggerated, leading to substantially increased ILDs as a sound source approaches a listener, thus making HRTFs distance dependent[13] [20][21][22]. This phenomenon makes the near-field the only space where a listener is able to estimate the distance of a sound source without any prior information about the intensity or spectrum of a source. Psychologically, the near-field is a very sensitive area to a listener, and thus important for personal situational awareness.

Due to the significant differences in localization cues between the near-field and the far-field, both spaces must be evaluated independently when assessing a Transparent Hearing System. The assessment should focus on perceptual testing, particularly on localization accuracy. Such assessment was beyond the scope of the current study.

7.2 Conclusions

1) Existing hear-through systems can badly disrupt the user's ability to localize sound. While this study has likely collected enough data to correlate specific product features to certain disruptions, more work is required to do such analysis. Further physical and behavioral tests should be performed to characterize the classes of these devices and fully rate them on a performance scale.

2) External devices such as helmets, goggles and muffs substantially alter all three localization cues (ITD, ILD and spectral characteristics), significantly deteriorating localization cues. While this study has collected data to correlate specific device details to localization cue deterioration, more work is required to do such analysis.

[13] In the near-field, the ITD's remain relatively constant with distance.

3) Customizing a Transparent Hearing System to the user's own ears appears to give a small improvement in elevation-localization performance over the use of generalized transfer functions. The value of this improvement, if confirmed in further testing, would have to be judged in relation to the cost of individualized HRTF measurements or customization.
4) If analytical work is to be done on the problem of transparent hearing, better performance metrics will be needed. Finding such metrics would require detailed research that attempts to find the relative psychoacoustic value of various stimulus features. There can be multiple redundant features, which are idiosyncratic to individual subjects, and there is no guarantee that subjects weigh different cues in the same way. In addition, the error metrics that appear to be most relevant are nonlinear functions' filter parameters, making for a difficult search problem.
5) A sound localization test protocol (or set thereof) is needed for military applications. The simple procedure developed here is a first step. Such a test protocol should measure subject performance while equipped with the Transparent Hearing System relative to their performance with open ears.
6) The flexibility of 32-channel system and its subset derivatives provides much more opportunity to investigate head-gear characteristics and explore optimization schemes.
7) Physical pinna/concha systems presented good localization cues, but with questionable aesthetics. Some approaches such as the hidden concha and integrated mechanical pinna deserve more study to potentially find an acceptable solution.
8) The simulated pinnae systems described here with delay-and-sum processing to generate elevation-dependent nulls showed a promising combination of performance and simplicity of processing. Eventual implementations could be self-contained, compact, analog devices. Further work should be devoted to advancing this approach.
9) The complexity of acoustically relevant head-gear is beyond the scope of the current numerical modeling methods.
10) Research is needed to determine the ability to adapt to localization cues altered by a transparent hearing system. Training was not a part of the current study. It is possible that the pattern of results could change if subjects were given long-term training.
11) This study did not examine several critical criteria, including: near/far-field effects, in-ear plugs, passive muffs, maximum system latency, and minimum direct path attenuation.

7.3 Future Work

This section summarizes the anticipated future work to complete the exploration, based on the findings of this work.

7.3.1 System Analysis

The most promising prototypes proposed in this document should be fully assessed using behavioral methods in order to determine their ability to support the localization of sound sources. The behavioral methods may include, but are not limited to:

1) Subjective localization performance
2) Minimum Audible Angle
3) Speech intelligibility
4) Adaptation
5) System Latency
6) Direct Path Attenuation

Subjective localization performance is aimed at measuring subjects' ability to accurately judge the perceived location of sound sources. When considering localization accuracy both near-field and far-field performance should be evaluated. A method similar to that used in [141] may be used to measure the performance for static heads. A test should also be performed to assess the accuracy of identifying the location of a sound source when listeners can employ all available strategies they choose during a localization task. Such method may be similar to that used in [81].

The Minimum Audible Angle (MAA) is an important measure that represents the sensitivity of a listener to spatial separation of sound sources as a function of the relative position of the source and listener. A method such as described in [128] may be used for this purpose.

Speech intelligibility should be measured using objective measures that may include the Speech Transmission Index (STI) or the Speech Intelligibility Index (SII) as described in [9][64][128].

For adaptation, it has been established in [63][141][146] that the human auditory system is able to adapt to foreign localization cues. Experiments conducted in [122][123] demonstrated the possibility of adaptation even to supernormal cues. The extent and speed of adaptation should be measured for the proposed systems.

Total system latency measurement and its effect on localization and task performance must be established. First, a physical assessment should be made of the total system latency. Second, perceptual testing should be conducted on human subjects to determine 1) whether such latency is noticeable; 2) its effect on user performance; and 3) users' adaptation capability to latency.

The direct path attenuation must be assessed using methods to measure the effective attenuation, as well as the perceived attenuation. Such testing will determine whether the direct path signal has been psycho-acoustically eliminated to support hear-through processing.

7.3.2 32-Channel Apparatus and General Array Systems

7.3.2.1 Direct HRTF

There are several variations on the general array systems that could improve results. The implemented 32-channel system used the physical location of the microphones to determine which HRTF filter to employ. However, a better result could be obtained by using the measured directivity patterns of the microphones to choose the HRTF filter. For example, the direction of peak response could be used as the direction for the HRTF. The peak can be found with the directivity centered at head center, or it can be centered at the measured physical location of the microphone. The head centered approach may yield better results for far-field sounds, while adjusting the selection basis to the physical microphone location may yield better near-field results. Additionally, improved results might be obtained by filtering with the minimum-phase HRTFs, since the time delay is already present due to the microphone location.

7.3.2.2 DTF to HRTF Filter Optimization

There are several variations on the filter optimization which could yield better results and should be investigated in the future.

Time Domain Optimization

When the desired filter length is small (less than 5 msecs), often a time-domain optimization will give a better result than using frequency-domain optimization with regularization. One method of time-domain optimization which should be investigated is the Mourjopoulos technique [47][102].

Minimum-Phase Least Squares

Since doing a full optimization on the maximum phase impulse responses in the complex plane does not yield good convergence for both magnitude and phase, better results might be obtained by doing a search based on the minimum phase versions of the respective filters. This method is similar to performing a magnitude-only search, but with the advantage of having a closed-form LSE solution.

Minimum-Phase Least Squares with Delay Search

Another variation on the Minimum Phase Least Squares would be to add a functional search for a single delay term to each filter. Since the HRTF can be represented well by a minimum-phase filter with a non-frequency dependent delay, an optimal set of filters might be found by searching for the minimum-phase filter and an optimum delay term separately.

7.3.2.3 DTF to Beams to HRTF Optimization

As implemented for this study, the set of beams were chosen to align with the HRTF filterset. However, since HRTF interpolation can be performed to generate any location, it might be advantageous to choose the set of beams that work best given the microphone layout and directivity. One possibility would be to calculate solutions to a large number of beams, and then choose a subset of those beams that meets some criteria based on quality of beam and spatial coverage.

7.3.2.4 Other Approaches

Subset search

For all optimization methods for general microphone array systems, it would be useful and enlightening to do a search of the possible subsets of microphones and determine the performance of the system. Conducting such a search would allow for an analysis of the channel count vs. performance curve and give some insight into which locations are most important.

High-Order Ambisonic

With the 32-channels as a capture device, it should be possible to optimize the system to produce an accurate set of high-order (2nd or 3rd order) B-Format signals. This spatial harmonic basis set can then be rendered to reproduce the sound field as heard by the listener.

7.3.3 Sound-Field Microphone Apparatus

There is future work to be done to fully implement the B-format encoding of the A-format microphone capsules and the creation of the B-format to binaural decoder. Standard free-field decoders will not apply because the head shadowing is already embedded in the signal. Therefore, a specialized encoder/decoder is needed to support the "head-shadowed binaural B-format" signal that is captured by this microphone array.

7.3.4 Simulated Pinnae Systems

The presented filter optimizations for simulated pinnae systems are dependent on a quality metric. These systems can be further improved by continued refinement of both the metrics and search algorithms that determine the filter parameters.

An analog system should be designed to study the viability of the suggested simple delay and sum two-microphone approach presented in 3.2.2.5.

7.3.5 Physical Pinnae Systems

Three distinct steps should be taken in future work relative to physical pinnae systems. First, a generic Transparent Hearing device on a muff-platform should be developed with integrated binaural microphone at the bottom of a pseudo ear canal. The platform should have a system for attaching interchangeable alternative physical shapes for the purpose of replicating pinna cues or replicating headgear accessories in close proximity to the hearing system. Secondly, a system of progressively changing the geometric shape on an interchangeable attachment will allow a study to iterate rapidly through many geometric alternatives. And finally, such a platform should be made wearable without tether so that users can readily train and potentially adapt to the foreign directional cues.

7.3.6 Numerical Modeling and Design

To complete the development of a usable HRTF computation package, several tasks remain, in addition to the numerical integration code. Software must be written to accept a surface triangulation produced by a geometric modeler (e.g. the visualization toolkit "VTK"), along with a beam direction and frequency, and call the integration routines in order to produce the matrix equation to be solved for the approximate solution to the Helmholtz PDE. This matrix equation must also be solved; by initially avoiding the more sophisticated variants of the BEM mentioned above, it should be possible to make simple use of Matlab's built-in GMRES routines for this problem. To handle all frequencies of interest in a practical amount of computer time, it would likely be necessary to move to one of these variants, however. Finally, the BEM actually produces a solution for the SPL field's amplitude and phase at all points on the surface of the scattering body; in addition to outputting this information for an input collection of microphone locations, it may also be useful to write software that allows visualization of these fields over the head/helmet surface to aid in selection of microphone positions.

7.3.7 General versus Custom HRTF's

Many systems discussed in this document utilize HRTFs measured on mannequins. Such generic data was shown to be less than optimal for localization and overall system performance. Because the use of personalized HRTFs enhances localization accuracy [140], future work should more deeply explore the effect of using personalized or adapted filter sets on 1) the speed of adaptability to a Transparent Hearing System; 2) localization accuracy; and 3) overall system performance.

7.3.8 Active Gain Control

The current implementations of Transparent Hearing System approaches did not demonstrate any active protection from harmful noise. Future work for the binaural and microphone array prototypes should include active gain control and/or active noise reduction in the signal path before the sound reaches the headphones and the listener's ears.

7.3.9 Signal Transmission Mechanism

Future work should include steps in determining the most practical and efficient signal transmission mechanism for a Transparent Hearing System. Issues to consider should include signal quality, localization cue retention, comfort, hygiene, and compliance with noise control.

7.3.10 Plugs vs. Muffs

The issue of the device type (or coupling to the listener) to use for noise control should be explored in future work. Earplugs, earmuffs, and their variants should be considered. Earplugs range from generic, passive devices to custom-molded hearing/communication plugs. Although earplugs provide good high-frequency hearing protection, their in-ear nature may prove to be impractical and unhygienic in harsh weather and/or combat conditions. Transducers located in earmuffs may be comfortable to wear for some tasks, but may result in less precise audio signal control due to interface variables.

A study should determine the most beneficial solution to the user per the determined operational task. Such studies should include evaluation of user enthusiasm, comfort, sound quality, hygiene, cost, and field-replacement.

7.3.11 Exploiting Microphone Arrays for Supernormal Performance

Hearing enhancement to include supernormal listening is a desired feature that has been the focus of recent research [38][114]. An important advantage microphone array systems have over binaural microphone systems is that they lend themselves well to supernormal listening capabilities. Future work should include the evaluation of supernormal performance derived from the microphones provided by the Transparent Hearing System.

7.3.12 Performance Metrics

Performance metrics are necessary and useful for 1) design and optimization of systems, and 2) for the validation and evaluation of systems. There are many dimensions to the effectiveness of a transparent hearing system. Several metrics have been proposed in this study, but they by no means span the range of system effectiveness. More metrics need to be explored and tested. Such metrics can be numerical or perceptual in basis. As suggested earlier, potentially a vector whose elements are a range of metrics could be devised and weighted to best describe overall system effectiveness.

8 Appendices

Appendix A: Microphone-Array Processing

Microphone-array processing is the form of signal processing through which the outputs of several microphones are filtered and combined in such a way as to create an overall system response that exhibits a directional-response characteristic (i.e., sources are amplified or attenuated based upon their location within the environment of the microphone array). This section provides a basic and brief introduction to microphone array systems and describes how they can be used to create a directional response. For a more complete description of these systems, please consult [64].

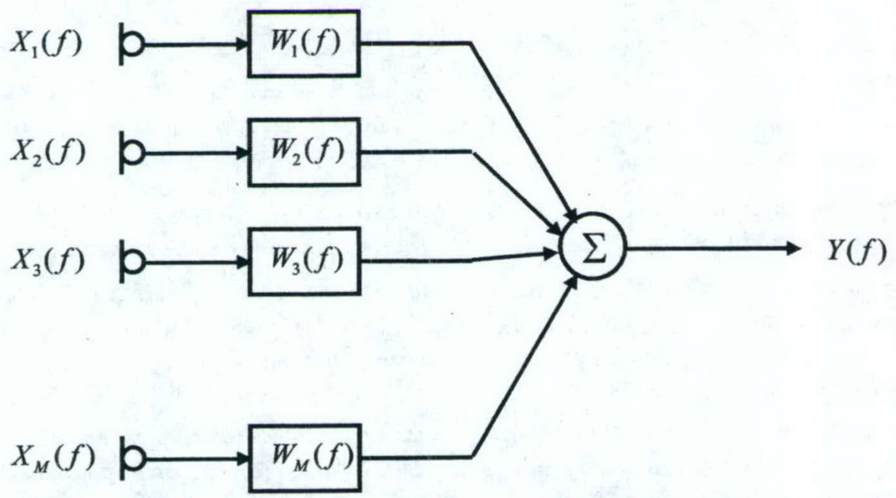


Figure 75: Diagram of a generic microphone array processing system. The output $Y(f)$ exhibits a directional characteristic described by the array filters, $W_m(f)$, and by the propagation properties from various source locations to the array microphones.

Figure 75 shows a basic M-microphone array system. This system generates an output signal $Y(f)$ by filtering the microphone signals $X_m(f)$ with filters $W_m(f), m = 1,2,\ldots,M$ and summing the results:

$$Y(f) = \sum_{m=1}^{M} X_m(f) W_m(f). \qquad (23)$$

Consider the arrival at the array microphones of a single signal $S(f,\rho,\theta,\phi)$ originating at the specific location (ρ,θ,ϕ). The microphone input components that arise due to $S(f,\rho,\theta,\phi)$ may be written as:

$$X_m(f) = H_m(f,\rho,\theta,\phi) S(f,\rho,\theta,\phi), \quad m = 1,2,\ldots,M, \qquad (24)$$

where the $H_m(f,\rho,\theta,\phi)$ is the source-to-microphone transfer function that describes the propagation of $S(f,\rho,\theta,\phi)$ from the source location to microphone m. This propagation encompasses factors such as travel-time, reverberation, sound scattering off of objects located near the microphone (e.g., a helmet), etc. Given these microphone inputs, then the array output due to $S(f,\rho,\theta,\phi)$ is:

$$\begin{aligned} Y(f) &= \sum_{m=1}^{M} W_m(f) X_m(f) = \sum_{m=1}^{M} W_m(f) H_m(f,\rho,\theta,\phi) S(f,\rho,\theta,\phi) \\ &= \underbrace{\left[\sum_{m=1}^{M} W_m(f) H_m(f,\rho,\theta,\phi)\right]}_{G(f,\rho,\theta,\phi)} S(f,\rho,\theta,\phi). \end{aligned} \tag{25}$$

The gain function $G(f,\rho,\theta,\phi)$ describes the effect of the array processing upon $S(f,\rho,\theta,\phi)$. This gain is directionally dependent due to the $H_m(f,\rho,\theta,\phi)$ terms – sources from different locations will propagate to the array microphones differently, and, consequently, they will experience different array gains. For this reason, $G(f,\rho,\theta,\phi)$ is also known as the directional response of the array.

The directional response of an array is governed by the source-to-microphone transfer functions $H_m(f,\rho,\theta,\phi)$ and by the array filters $W_m(f)$. Array processing systems with specific directional response characteristics are designed by manipulating these properties. The $H_m(f,\rho,\theta,\phi)$ are largely determined by the array environment (reverberation, source scattering, etc.), but some control of these responses is possible through the choice of the array configuration: as the microphone placements vary, the $H_m(f,\rho,\theta,\phi)$ also vary. The array filters $W_m(f)$, on the other hand, are under complete user control and are selected in a variety of ways depending upon the desired application. For example, the $W_m(f)$ might be selected so that the resulting $G(f,\rho,\theta,\phi)$ is a least-squares approximation to a desired directional response. Alternatively, the $W_m(f)$ might be selected or even continually adapted to yield maximal sensitivity to sources from one particular location while attenuating all other sources [134]. One final possibility of choosing $W_m(f)$ arises when there are only L sources in the environment and there are fewer sources than microphones ($L \leq M$). In this case, if the $H_m(f,\rho,\theta,\phi)$ for the individual sources are known, then it is possible to choose L sets of filters $W_{l,m}(f)$ that yield L directional responses $G_l(f,\rho,\theta,\phi)$, $l = 1,2,\ldots,L$, that can extract each of the component sources individually. The main issue with this approach is obtaining knowledge of the $H_m(f,\rho,\theta,\phi)$ for each source in the acoustic environment, and some current methods estimate them using knowledge of microphone-array geometry and sound-propagation models (the best known of these methods is MUSIC [115]). Another, more recent approach called Independent Component Analysis (ICA), uses signal statistics to estimate the $H_m(f,\rho,\theta,\phi)$ [13]. Specifically, it tries to find the maximally-independent set of sources that result in the observed source mixtures received by the microphones. While this approach is very interesting with great potential (including, perhaps, the elimination of the requirement that there be fewer sources than microphones), it remains an area of active research and is not yet practical for implementation at this time. For this reason, this work concentrates on more traditional microphone-array approaches to acoustic transparency.

Appendix B: Ambisonics

Ambisonics is a surround sound system developed in the 1970's by Michael Gerzon [89]. It is based on a mathematical model of directional psychoacoustics and is capable of capturing, transmitting and reproducing a three-dimensional sound field. Unlike 5.1 and other surround systems, the transmitted signals do not correspond to direct speaker feeds. Instead, the transmitted signals correspond to spatially orthogonal pressure signals that can be decoded to any size speaker array. There are several advantages to Ambisonics over traditional surround mixing: it takes into account more directional cues; it has good inter-loudspeaker imaging which leads to improved image stability; the sound-field can be rotated; and the decoding can be precisely tuned to the individual listening environment [98]. In addition, there exist commercially available Ambisonic microphones that can directly and accurately capture the three-dimensional sound-field [126]. In addition to use in recording studios, Ambisonic-based sound-field microphones have been used at NASA for 3D analysis of sound fields [58].

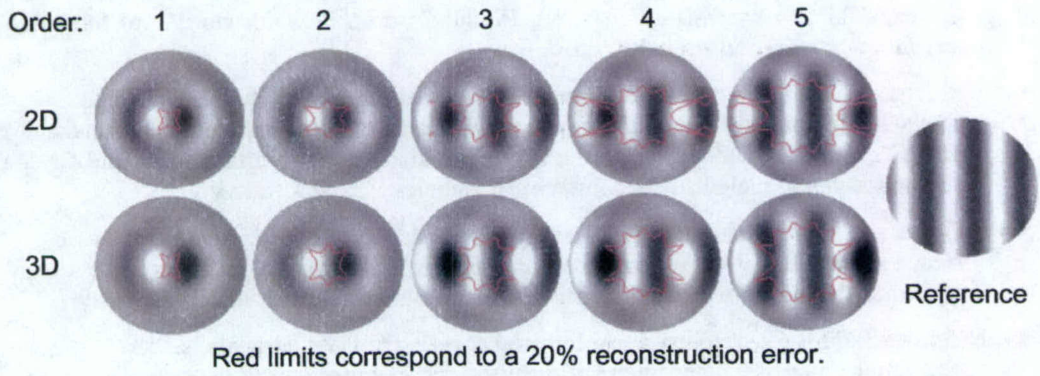


Figure 76: Spectral Reconstruction of acoustic field of plane wave [31]

The core of Ambisonic theory is based around representing the sound field at a point by decomposing the pressure into spherical harmonics (see Figure 76 and Figure 15). The 0^{th} order harmonic is strictly the omni-directional pressure at the center of the sound-field. This signal is referred to as W and is identical to what an ideal omni-directional free-field microphone would produce. The 1^{st} order harmonics correspond to the pressure gradient, (the first partial derivative in each coordinate direction) which is proportional to particle velocity [28]. These signals are called X, Y, and Z, and correspond to three figure-eight microphones oriented along the coordinate axes. The higher order harmonics correspond to higher order derivatives. Thus, Ambisonic theory can be thought of as three-dimensional Taylor Series approximation to the sound field at a point [31][106][107]. By creating the pressure and first derivative correctly, then the sound field will be approximately correct for a region around the center point. (See Figure 15) This allows the sound field to be physically reproduced at the ears when the head is in the middle of the sound-field, and allows for natural perception of the three-dimensional sound-field.

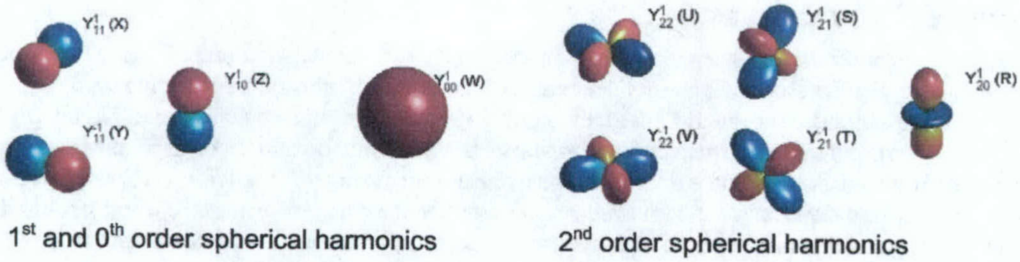


Figure 77: Ambisonic theory spherical harmonics.

There are several different signal formats that are in common use for Ambisonics [11][42]:

- A-format: This is the raw data from a sound-field microphone, which consists of 4 cardioid microphones arranged in a tetrahedron.
- B-format: This is the fundamental representation of Ambisonics, with each of signal corresponding directly to one of the spherical harmonics. Most commonly B-format consists of W, X, Y and Z. The Z can be left out to give a horizontal two-dimensional sound-field. In addition, it can also include the second order harmonics R, S, T, U and V.
- C-format: Also referred to as (HUJ). This is a specification of matrix versions of the B-format signals for delivering signals to the consumer. The primary goal was to get a set of signals that would be backward compatible for conventional stereo and mono display.
- D-format: This is the decoded signal format that is sent to the loudspeakers. The specification depends on the number of speakers and their positions in the rooms.
- G-format: This is a 5.1 compatible decoding of B-Format. Essentially, it can be thought of as a D-format decoding for speakers that are in the positions for a 5.1 setup. This format can be made to be reversible so that the original B-format signal can be recovered and then re-decoded to allow for different room configurations.

Ambisonics can be easily split into two independent halves, encoding and decoding. Encoding is the process of decomposing the sound into the spherical harmonics and encoding them into B-format. Sounds can be recorded directionally in 3-D by using a sound-field microphone to record into A-format and then converting the resulting signals into B-format. Alternatively, normal monophonic sounds can be converted into B-format by scaling the signal by the response of a given spherical-harmonic in the direction of the simulated source. Simulated and recorded sound-fields can be mixed to give a total sound environment.

Signal decoding is the process of taking the B-format signal and delivering it to the listener's ears. Most often this is accomplished through a regular speaker array. The more speakers the better, but acceptable results can be achieved with 4 speakers in a square for a two-dimensional sound-field and 8 speakers in a cube to display elevation information as well. The decoding stage is also where additional transformations can be applied, such as rotating the sound-field and adjusting the "presence" and "dominance" [11]. Alternatively, the signal can be transcoded into a binaural mix for headphone listening also called "Binaural B-format" [68][66][80]. Headphone rendering has the advantage that the location of the speakers relative to the ears is known, so that there is no problem with having a "sweet spot". In addition, since B-format signals can be rotated prior to decoding, it is possible to use head-tracking to preserve the orientation of the sound-field [88][132].

Appendix C: Audio System Characterization

The characterization of an acoustic system is performed by measuring the system's impulse response. During most acoustic impulse response measurements, the system under test is assumed to be linear and time invariant, that is, a *linear time-invariant* (LTI) system. In a time invariant system, the fundamental properties do not change with respect to time. In a linear system, the response characteristics are additive: the output of a sum of inputs is equal to the sum of outputs produced by each input individually. The relationship between the input and output of an LTI system can be expressed by the response of the system in either the time or frequency domain by:

$$y(t) = h(t) * x(t)$$
$$Y(t) = H(t)X(t) \qquad (26)$$

where '*x*' and '*X*' indicate the input signal, '*y*' and '*Y*' the output, '*h*' and '*H*' the response function of the system, and * denotes convolution.

The extraction of the response of the system is performed by cross-correlating the input signal with the resulting output, thereby deconvolving (◊) the two sequences. From ($Y(t) = H(t)X(t)$ (26) the system response is defined as:

$$h(t) = y(t) \lozenge x(t)$$
$$H(t) = Y(t)/X(t) \qquad (27)$$

Thus, the system can be described by a *frequency response function* $H(f)$, which is defined as the Fourier Transform of the impulse response function $h(t)$:

$$H(f)= \int_{0}^{\infty} h(t)e^{-j\omega t}dt \qquad (28)$$

Numerous excitation signals can be used for impulse response measurements, including pure tones, noise bursts, and pseudo-random noise sequences. Choosing the appropriate test signal largely depends on the measurement circumstances, including the reproduction and recording equipment, as well as the acoustic environment in which the measurements are being taken. The excitation signal should, ideally, have a perfectly flat frequency spectrum.

Two of the most popular pseudo-random noise test signals used today are the *maximum-length sequence* (MLS) [111] and the *Golay code pair* [51][56][145][147]. The *MLS* is a deterministic sequence of integers. It can be produced by three-stage shift registers using the *exclusive-or* operation. The MLS is a binary sequence, resulting in integers +1 and -1, and has a length of $2^N - 1$. The stimulus has an evenly distributed energy and, similarly to the Golay codes, the MLS has a flat frequency spectrum, and random phase.

The Golay codes are two binary sequences that have complementary frequency spectra, that is, the sum of the auto-correlation of the sequences results in a perfectly flat power spectrum. Although Golay codes can be of any length, most algorithms construct Golay codes whose length is exactly a power of two. Golay codes have a superior signal-to-noise ratio (SNR) when compared to the MLS: the SNR increases by 3dB with every doubling of signal length and is defined by $dB = 10\log_{10}(2L)$ where L is the length of the Golay sequence.

A complementary Golay sequence can be constructed by a "negate and concatenate" algorithm defined as

$$a = [a\ b],\ \ b = [a\ –b] \qquad (29)$$

Starting with the pair *a* and *b*, a Golay code sequence of length 2^L may be generated by recursively applying ($a = [a\ b],\ b = [a\ –b]$ (29).

The captured impulse response not only contains the spectral characteristics of the measured acoustic system. Time delays, noise, and characteristics of reproduction and recording devices, as well as other electronics, are included in the impulse response. The recorded response must be processed in order to extract the true characteristic of the measured system. The processing of raw (recorded) impulse responses may include any or all of the following: direct sound extraction, time delay estimation, and measurement system compensation (equalization).

To extract the direct sound of the measurement and discard any unwanted reflections, the measured impulse responses are windowed using a rectangular, or other type, window. Each measured channel should be windowed individually using the same window for all channels.

The direct path impulse response is correlated with its minimum phase equivalent to estimate its starting time within the direct path window. This starting time is added to the direct path window starting time to determine the speaker-microphone travel time. To calculate inter-channel time delay, the difference in the travel time for each channel is calculated. For example, in HRTF measurement: the left-ear and right-ear travel times are differenced to estimate interaural time delay, and averaged and scaled by the speed of sound to estimate speaker-subject range.

The measured impulse response does not only contain the head-related impulse response. The response of the equipment used for the measurement is included in the measurement. The equipment characteristics include the speaker and microphone frequency response, A/DD/A, sound card, speaker amplifier and microphone pre-amplifier responses. In order to extract the pure filters, responses must be compensated for the measurement equipment.

The measured transfer function can be defined as:

$$Y_s(\omega) = X_s(\omega)\, S(\omega)\, M(\omega)\, H(\omega) \qquad (30)$$

where $X_s(\omega)$ is the test signal, $S(\omega)$ is the transfer function of the loudspeaker and amplifier, $M(\omega)$ is the transfer function of the microphone and pre-amplifier, and $H(\omega)$ is the head-related transfer function.

The free-field transfer function is used as the compensation transfer function to the system, defined as:

$$Y_{comp}(\omega) = X_s(\omega)\, S(\omega)\, M(\omega) \qquad (31)$$

The transfer function for the measurement apparatus can be measured with precision sound calibration equipment. The inverse transfer function is used to equalize the measurements. The HRTF ($H(\omega)$) is obtained by removing the free-field transfer function from the measured response:

$$H(\omega) = Y_s(\omega) \,/\, Y_{comp}(\omega) \qquad (32)$$

For a more generalized data set of HRTFs, *free-field* and *diffuse-field* equalization can be considered. Free-field equalization is obtained for each ear by dividing the data set by a reference measurement, typically the response of the system when microphones are positioned in the free-field, however a reference location may also be used (e.g. frontal location at 0° azimuth and 0° elevation). Diffuse-field equalization is derived by the power of the transfer function of measured HRTFs. A diffuse-field transfer function is obtained by power averaging all HRTFs from each ear, and taking the square root of the averaged power. Equalized HRTFs are obtained by dividing the original measurement by the diffuse-field HRTF of that ear.

$$H_{df} = \frac{H(\omega,\theta,\varphi)}{\sqrt{\frac{1}{N}\sum_{\theta i}\sum_{\varphi j}\left|H(\omega,\theta_i,\varphi_j)\right|}} \qquad (33)$$

This results in the removal of all common characteristics to the measurements, i.e. not incident-dependent factors such as reproduction and recording equipment.

HeadZap is a commercially available HRTF measurement system developed by AuSIM. It is designed to operate in reflective, noisy settings typical of offices and laboratories. HeadZap uses one or multiple loudspeakers mounted on adjustable arms. HeadZap's measurement and processing methods combine to remove the effects of reflections, increase measurement signal-to-noise ratio, and account for subject positioning errors.

The subject is seated on a rotating stool, outfitted with blocked meatus microphones and may also be equipped with a head-tracking device to monitor the position and movements of the subject. Golay codes are used as the test signal. A graphical user interface allows the user to set the measurement parameters including locations to be measured, golay code length, impulse response length, sampling rate (44.1kHz, 48kHz, or 96kHz), field equalization (diffuse or free-field), and other compensation parameters.

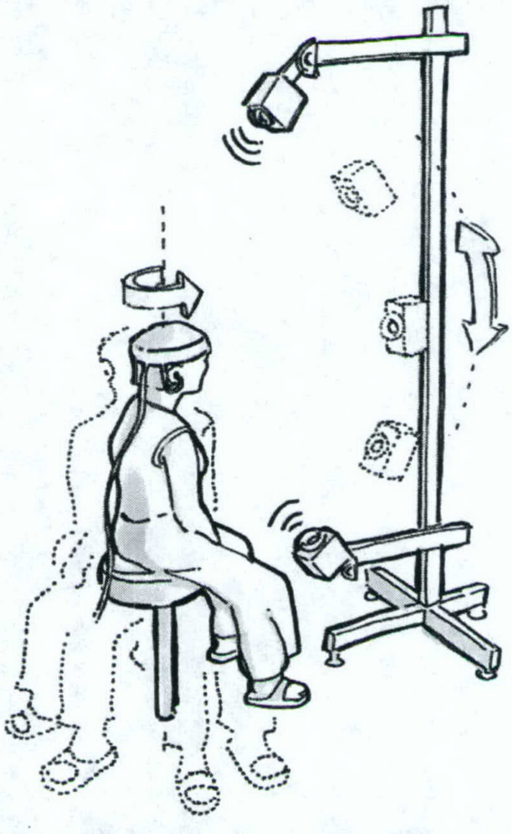


Figure 78: HeadZap apparatus operation, showing the two degrees of freedom: 1) one or more loud-speaker positioned along an arc, and 2) the subject turning to a select number of bearings.

All HRTFs measured by HeadZap are stored as an Acoustic Head Map and are immediately available for rendering on the AuSIM3D renderer.

HeadZap is bundled with AuProbe, a flexible system identification utility capable of measuring audio-band electronic and acoustic systems and devices, including the reflection and transmission properties of materials and objects. Test signals are passed directly into AuProbe as an array of numbers; therefore *any* signal can be constructed by the operator and used as the test signal. The I/O of AuProbe is sample-accurate and synchronized, making AuProbe an ideal tool not only to measure spectral characteristics of acoustic systems, but also propagation and inter-channel delays.

THIS PAGE IS INTENTIONALLY BLANK.

Appendix D: Device Data

An immense amount of data was collected during this exploration. Follow-on work should begin with an analysis of this data.

The table across the following two pages (112 and 113) describes the datasets collected and available for analysis.

Six pages (114 through 119) follow the table depicting many of the configurations tested.

Three pages following the pictorial (120 through 122) present a sample datasheet of one particular device configuration demonstrating the amount of information available for each device.

For interested parties, the datasets may be obtained from the authors. The datasets may be experienced aurally through AuSIM's audio simulation systems. A particularly useful tool that may be supplied with the datasets is the AuSIM Rendograph™ application. Rendograph loads a dataset for auralization, presenting the listener with an interactive graphic of the spherical grid points sampled. The listener can select a test signal and while listening over headphones, the listener hears that signal as if they were wearing the device measured. An advanced version of Rendograph allows the loading of multiple datasets for direct comparison.

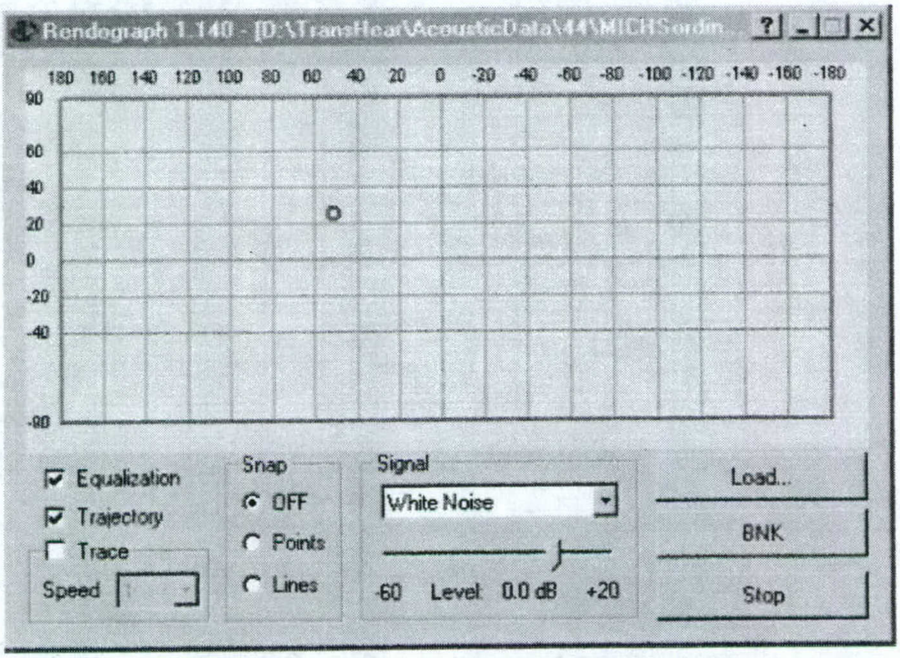
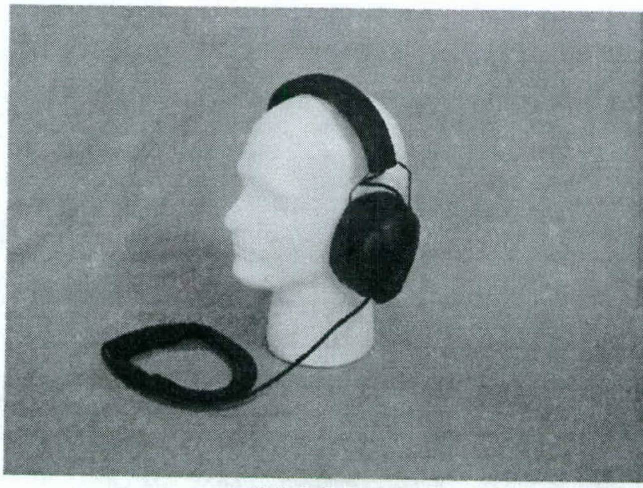


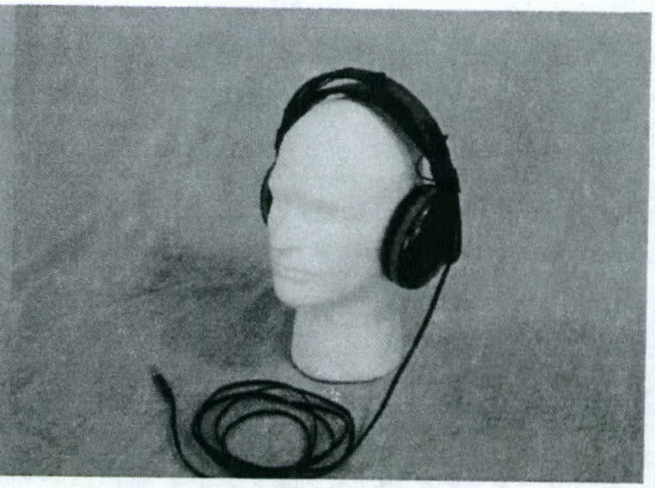
Figure 79: Rendograph applet screenshot. The spherical map is presented in direct polar projection, and thus the top and bottom lines are each one polar point (straight-up and straight-down respectively). The intersection of gridlines depict locations actually measured. All points in between are interpolated at render time. The blue-dot depicts the currently rendered spherical location, which can be moved interactively with the mouse.

Reference Name	File	Description	Measurement Facility
BNK Reference			
BNK	BNK.ahm	Bruel & Kjaer HATS mannequin	NASA Ames
Hear-Thru Devices			
MICH_Sordin	MICHSordinsActive.ahm	TC2000 helmet with Sordin Supreme III HT ON	NASA Ames
SordinSupreme	SordinActive.ahm	Sordin Supreme III HT ON	NASA Ames
Leightning	LeightningActive.ahm	Howard Leight Leightning HT ON	NASA Ames
Bilsom707Impact	Bilsom707Active.ahm	Bilson 707 Impact II HT ON	NASA Ames
Radians	RadiansActive.ahm	Radians Pro-Amp Electronic Earmuffs HT ON	NASA Ames
Remington	Remington2000Active.ahm	Remington R2000 HT ON	NASA Ames
Leakage Measurements			
HD205	Leakage_SennHD205.ahm	Sennheiser HD205	NASA Ames
HD250	Leakage_sennHD250.ahm	Senheiser HD250	NASA Ames
HD540	Leakage_SennHD540.ahm	Sennheiser HD540	NASA Ames
HDC200	Leakage_SennHDC200ANR.ahm	Sennheiser HDC200 ANR ON	NASA Ames
HDC200	Leakage_SennHDC200noANR.ahm	Sennheiser HDC200 ANR OFF	NASA Ames
SilencioMuffs	Silencio.ahm	Silencio Low-Pro 2000 Passive Hearing Protectors	NASA Ames
MICH_Sordin	MICHSordinsHTOff.ahm	TC2000 helmet with Sordin Supreme III HT OFF	NASA Ames
SordinSupreme	SordinHTOff.ahm	Sordin Supreme III HT OFF	NASA Ames
Leightning	LeightningHTOff.ahm	Howard Leight Leightning HT OFF	NASA Ames
Bilsom707Impact	Bilsom707HTOff.ahm	Bilson 707 Impact II HT OFF	NASA Ames
Radians	RadiansHTOff.ahm	Radians Pro-Amp Electronic Earmuffs HT OFF	NASA Ames
Remington	Remington2000HTOff.ahm	Remington R2000 HT OFF	NASA Ames
MICH Helmet with Accessories			
MICH	MICH.ahm	CGF/Gallet TC2000 standard, basis for MICH	NASA Ames
MICH_JSLIST	MICHChemBio.ahm	TC2000 + JSLIST mask	NASA Ames
MICH_Goggles	MICHGogglesMuffs.ahm	CGF/Gallet TC2001 "side-cut" with Sennheiser HME100	NASA Ames
MICH_NVG	MICHNVGdown.ahm	TC2000 with NightVisionGoggles down	NASA Ames
Scorpion R2 Helmet with Accessories			
R2_VisorSansMuffs	ScorpionNoMuffs.ahm	Scorpion R2, accessories mounted over ears	NASA Ames
R2_VisorMuffsAcc	ScorpionMuffs.ahm	Scorpion R2, accessories and muffs mounted over ears	NASA Ames
R2_SansMuffsAcc	ScorpionNoMuffsNoAttch.ahm	Scorpion R2, no accessories	NASA Ames
Other Helmets and Accessories			
PASGT_Goggles	PASGTGoggles	Personal Armor System, Ground Troops helmet	NASA Ames
NewChemBio	ChemBioMask.ahm	Prototype chem-bio mask	NASA Ames
JSLISTXS	ChemBioJSLIST	Standard chem-bio mask issue JSLIST/XS	NASA Ames
KEMAR Reference			
KEMAR_MIT	KEMAR_MIT.ahm	Gardner and Martin's 1994 KEMAR data from MIT	MIT Media Lab
KEMAR_Sens	KEMAR_Sens.ahm	Sensimetrics' KEMAR 2003	Sensimetrics
TC2001_KEMAR	TC2001KEMAR.ahm	Sensimetrics' KEMAR 2003 with TC2001 SideCut	Sensimetrics
Hear-Thru Devices			
Peltor	PeltorCOMTACActive.ahm	Peltor COMTAC HT On	Sensimetrics
HiddenConcha	HiddenConcha.ahm	KEMAR Concha embedded in HD205 with TC2001	Sensimetrics
SimPinna_LSM_1-A	SimPinna1aLSM.ahm	1-microphone Simulated Pinna on TC2001/HD205	Sensimetrics
SimPinna_LSM_1-B	SimPinna1bLSM.ahm	1-microphone Simulated Pinna on TC2001/HD205	Sensimetrics
SimPinna_LSM_2-A	SimPinna2aLSM.ahm	2-microphone Simulated Pinna on TC2001/HD205	Sensimetrics
SimPinna_LSM_2-B	SimPinna2bLSM.ahm	2-microphone Simulated Pinna on TC2001/HD205	Sensimetrics
SimPinna_DEL_2-A	SimPinna2aDEL.ahm	2-microphone Simulated Pinna on TC2001/HD205	Sensimetrics
SimPinna_DEL_2-A	SimPinna2bDEL.ahm	2-microphone Simulated Pinna on TC2001/HD205	Sensimetrics
SimPinna_LSM_4	SimPinna4aLSM.ahm	4-microphone Simulated Pinna on TC2001/HD205	Sensimetrics
SimPinna_LSE_8	GenArray8LSE.ahm	8-microphone General Array on TC2001/HD205	Sensimetrics
SimPinna_LSE_14	GenArray14LSE.ahm	14-microphone General Array on TC2001/HD205	Sensimetrics
ARO reference			
ARO	ARO.ahm	ARO subject	Satellite Studios
Hear-Thru Devices			
TC2001_24-8	aro-32-ch-direct-hrtf.ahm	32-channel system	Satellite Studios
TC2001_24-8	aro-32-ch-opt-bnk48.ahm	32-channel system with optimization	Satellite Studios
PinnaeSoftMuff	SoftMuffPinna.ahm	Human-like Soft Pinnae on HD205	Satellite Studios
PinnaeHardMuff	HardMuffPinna.ahm	Machined Hard Pinnae on HD205	Satellite Studios
HelmetConcha_R2	ConchaInR2Muff.ahm	Concha in R2 muff	Satellite Studios
HelmetConcha_R2	ConchaInR2Helmet.ahm	Concha on R2 helmet	Satellite Studios
Leakage Measurements			
PinnaeHardMuff	Leakage_HardMuffPinna.ahm	Machined Hard Pinnae on HD205	Satellite Studios
Other Helmets and Accessories			
PASGT_Goggles	PASGTGoggles.ahm	PASGT with Goggles	Satellite Studios
MICH_Goggles	MICHGoggles.ahm	MICH with goggles	Satellite Studios

Reference Name	Sampling rate	Filter Length	Elevation (degrees)			Azimuth (degrees)			Locations measured	Helmet	Muffs	Goggles
			Origin	Interval	Arc	Origin	Interval	Arc				
BNK Reference												
BNK	96kHz	256	70	-10	110	-180	10	360	432	none	none	none
Hear-Thru Devices												
MICH_Sordin	96kHz	256	70	-10	110	-180	10	360	432	MICH	Supreme III	none
SordinSupreme	96kHz	256	70	-10	110	-180	10	360	432	none	Supreme III	none
Leightning	96kHz	256	60	-30	90	-180	20	360	72	none	Leightning	none
Bilsom707Impact	96kHz	256	60	-30	90	-180	20	360	72	none	707 Impact II	none
Radians	96kHz	256	60	-30	90	-180	20	360	72	none	Pro-Amp	none
Remington	96kHz	256	60	-30	90	-180	20	360	72	none	R2000	none
Leakage Measuremen												
HD205	96kHz	256	70	-10	110	-180	10	360	432	none	HD205	none
HD250	96kHz	256	70	-10	110	-180	10	360	432	none	HD250	none
HD540	96kHz	256	70	-10	110	-180	10	360	432	none	HD540	none
HDC200	96kHz	256	70	-10	110	-180	10	360	432	none	HDC200	none
HDC200	96kHz	256	70	-10	110	-180	10	360	432	none	HDC200	none
SilencioMuffs	96kHz	256	60	-30	90	-180	20	360	72	none	Low-Pro 2000	none
MICH_Sordin	96kHz	256	70	-10	110	-180	10	360	432	MICH	Supreme III	none
SordinSupreme	96kHz	256	70	-10	110	-180	10	360	432	none	Supreme III	none
Leightning	96kHz	256	60	-30	90	-180	20	360	72	none	Leightning	none
Bilsom707Impact	96kHz	256	60	-30	90	-180	20	360	72	none	707 Impact II	none
Radians	96kHz	256	60	-30	90	-180	20	360	72	none	Pro-Amp	none
Remington	96kHz	256	60	-30	90	-180	20	360	72	none	R2000	none
MICH Helmet with Ac												
MICH	96kHz	256	70	-10	110	-180	10	360	432	MICH	none	none
MICH_JSLIST	96kHz	256	70	-10	110	-180	10	360	432	MICH	none	none
MICH_Goggles	96kHz	256	70	-10	110	-180	10	360	432	SideCut	HME100	(orange)
MICH_NVG	96kHz	256	70	-10	110	-180	10	360	432	MICH	none	none
Scorpion R2 Helmet w												
R2_VisorSansMuffs	96kHz	256	70	-10	110	-180	10	360	432	R2	none	built-in
R2_VisorMuffsAcc	96kHz	256	70	-10	110	-180	10	360	432	R2	R2 integrated	built-in
R2_SansMuffsAcc	96kHz	256	70	-10	110	-180	10	360	432	R2	none	built-in
Other Helmets and Ac												
PASGT_Goggles	96kHz	256	70	-10	110	-180	10	360	432	PASGT	none	yes
NewChemBio	96kHz	256	70	-10	110	-180	10	360	432	none	none	none
JSLISTXS	96kHz	256	70	-10	110	-180	10	360	432	none	none	none
KEMAR Reference												
KEMAR_MIT	44.1kHz	512	-40	10	130	-180	5	360	710	none	none	none
KEMAR_Sens	48kHz	128	60	-30	90	-180	-30	360	48	none	none	none
TC2001_KEMAR	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	none	none
Hear-Thru Devices												
Peltor	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	COMTAC	none
HiddenConcha	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_LSM_1-A	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_LSM_1-B	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_LSM_2-A	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_LSM_2-B	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_DEL_2-A	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_DEL_2-A	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_LSM_4	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_LSE_8	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
SimPinna_LSE_14	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
ARO reference												
ARO	48kHz	128	60	-30	90	-180	-30	360	48	none	none	none
Hear-Thru Devices												
TC2001_24-8	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
TC2001_24-8	48kHz	128	60	-30	90	-180	-30	360	48	SideCut	HD205	none
PinnaeSoftMuff	48kHz	128	60	-30	90	-180	-30	360	48	none	HD205	none
PinnaeHardMuff	48kHz	128	60	-30	90	-180	-30	360	48	none	HD205	none
HelmetConcha_R2	48kHz	128	60	-30	90	-180	-30	360	48	R2	R2	none
HelmetConcha_R2	48kHz	128	60	-30	90	-180	-30	360	48	R2	none	none
Leakage Measuremen												
PinnaeHardMuff	48kHz	128	60	-60	60	0	0	0	2	none	HD205	none
Other Helmets and Ac												
PASGT_Goggles	48kHz	128	60	-30	90	-180	-30	360	48	PASGT	none	yes
MICH_Goggles	48kHz	128	60	-30	90	-180	-30	360	48	MICH	none	yes



HD205

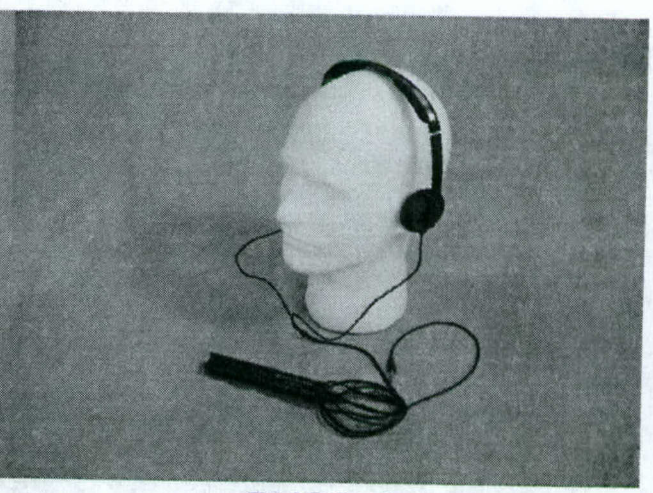


HD250

HD540

HDC200

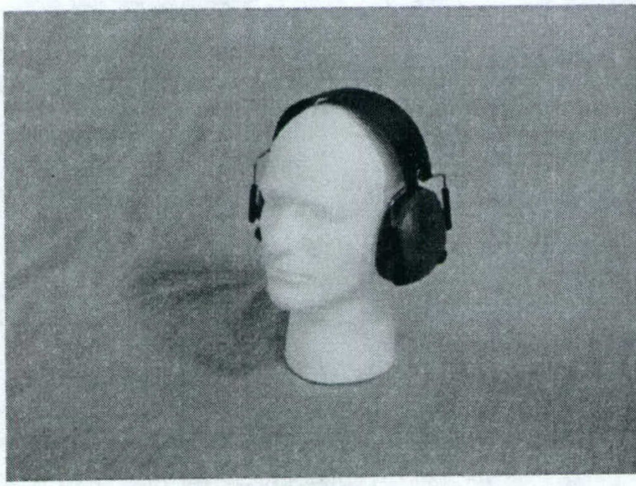
HME100



PXC250

Leightning

Bilson707Impact

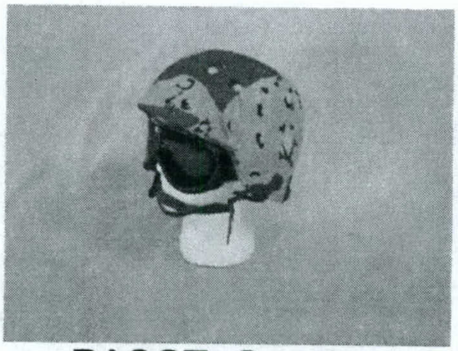


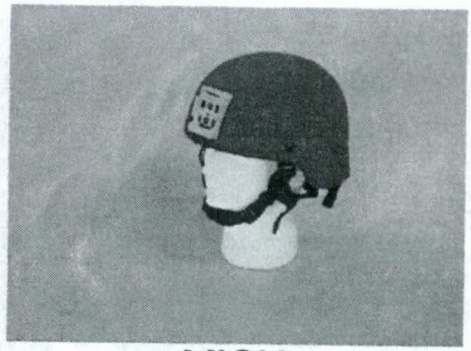
Radians

Remington

SilencioMuffs

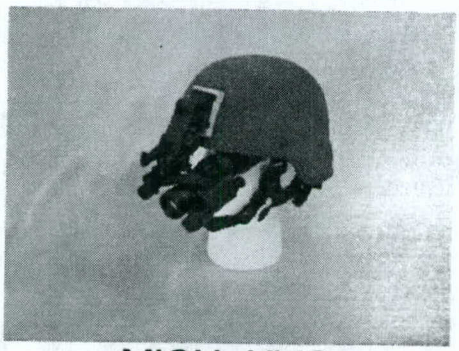
SordinSupreme


PASGT_Goggles

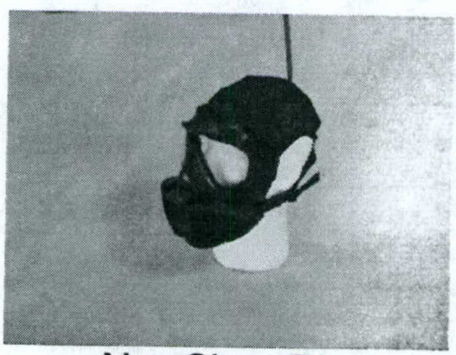

MICH

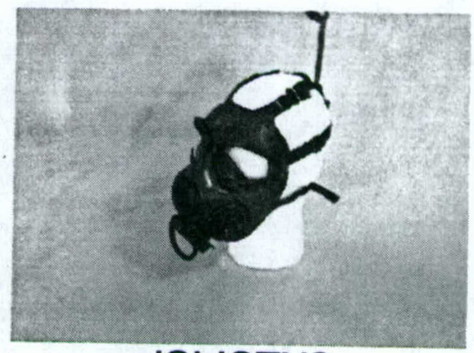

MICH_Goggles

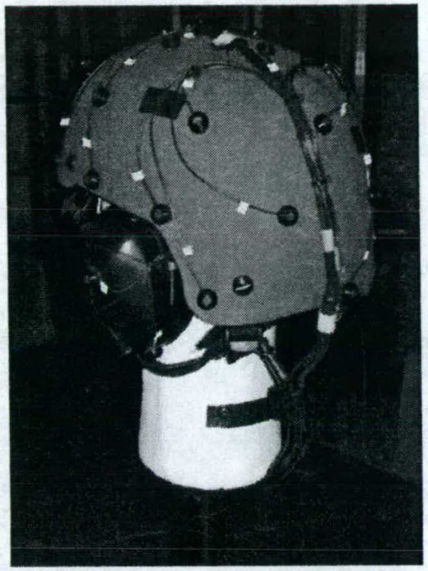

MICH_Sordin


MICH_NVG

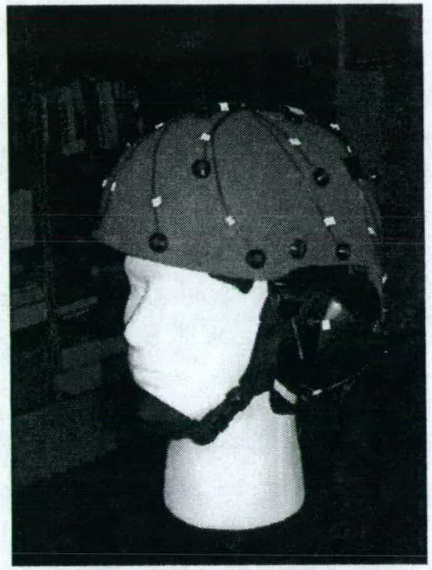

MICH_JSLIST


NewChemBio

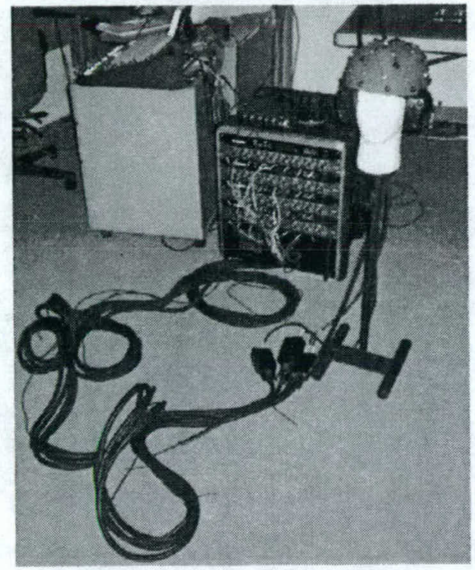

JSLISTXS



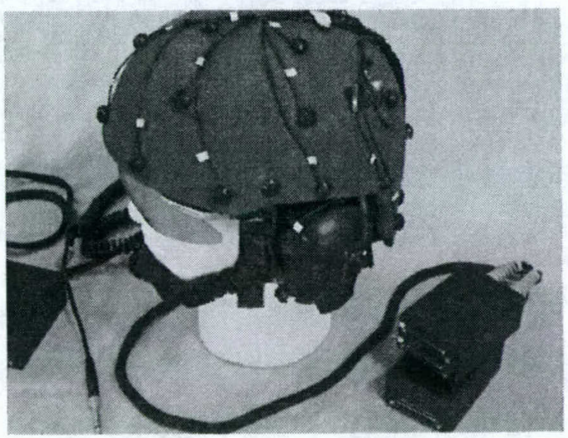
TC2001_24-8_back



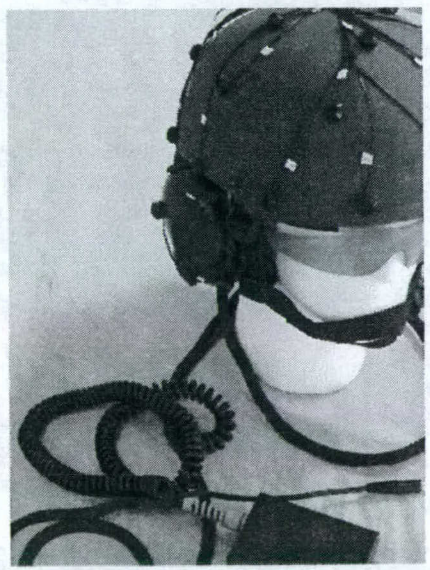
TC2001_24-8_front



32ChnlTestSystem



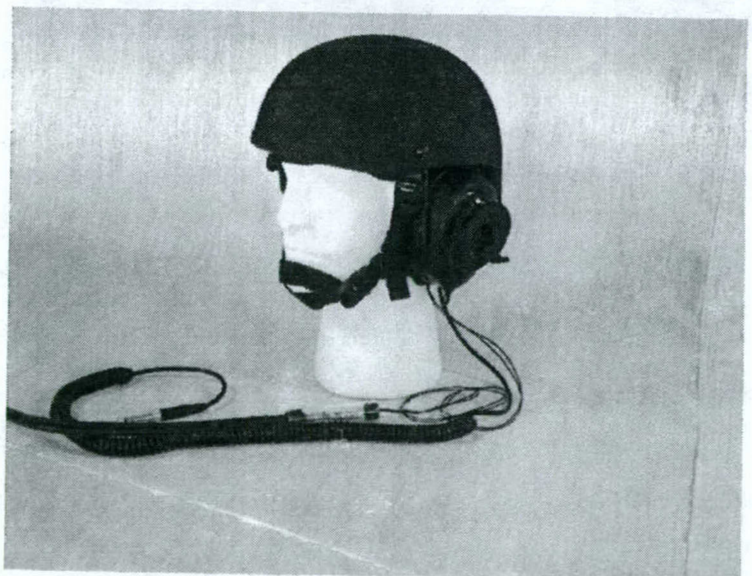
TC2001_24-8_left



TC2001_24-8_right


TC2001_HME100


TC2001_SoftMuff

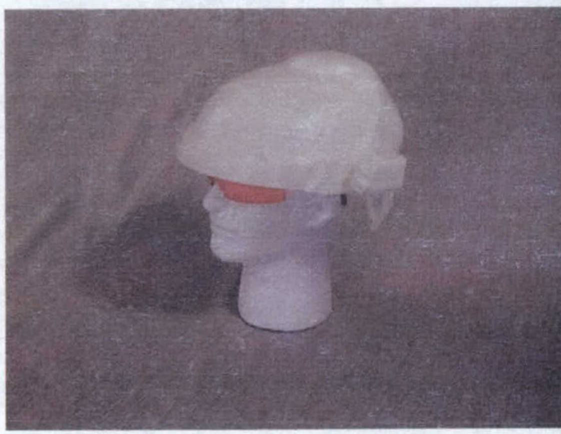

TC2001_HardMuff


TC2001_SndFldMuff

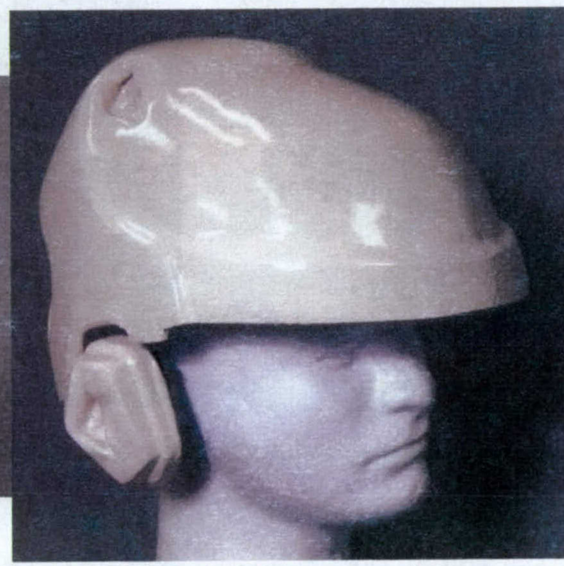
PinnaeHardMuff

PinnaeSoftMuff

SoundFieldMuff



R2_VisorSansMuffs



HelmetConcha_R2

InEarPassive

SORDIN SUPREME III HEARING PROTECTOR with CGF/Gallet TC-2000

Supreme III Earmuffs

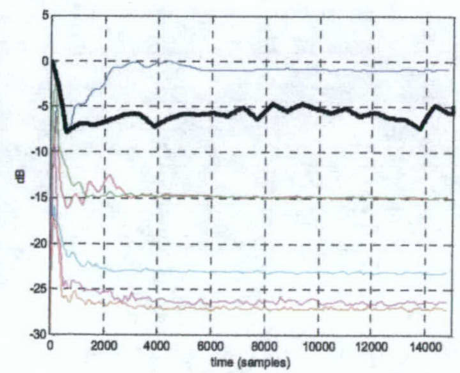
- Level-dependent
- Designed to conform to international mil-specs and extreme conditions
- Microprocessor-controlled
- Low current consumption
- Foldable

Headband

SNR = 28 dB H = 29 dB, M = 26 dB, L = 18 dB

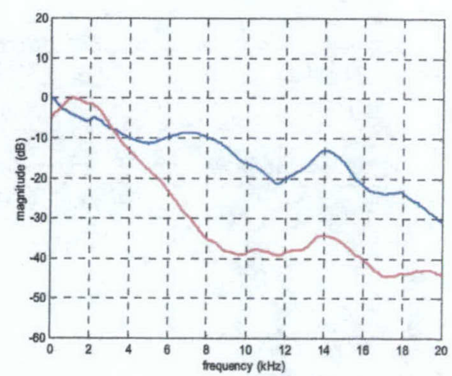
Frequency, f (Hz)	63	125	250	500	1000	2000	4000	8000
Mean value, mf (dB)	15.1	14.2	20.0	28.7	32.5	28.1	38.4	38.1
Std. Deviation, sf (dB)	3.3	2.9	3.0	3.5	2.2	2.5	2.2	4.3
APV, mf-sf (dB)	11.8	11.3	17.0	25.2	25.6	25.6	36.2	33.8

AGC-t



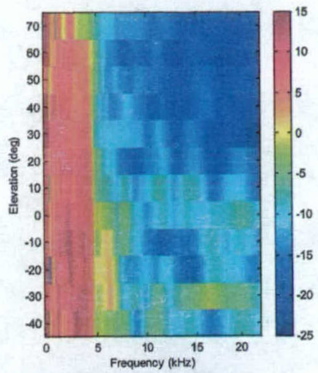
System AGC onset curve: black (bold), 250Hz (blue), 500Hz (red), 1kHz (green), 2kHz (magenta), 4kHz (cyan), 8kHz (orange)

AGC-f

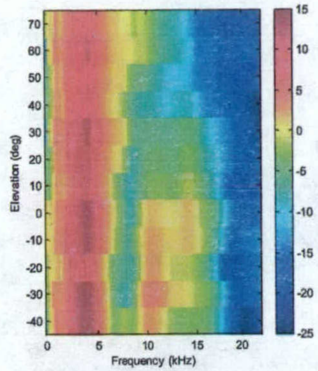


Frequency response of system with AGC active (red), and inactive (blue)

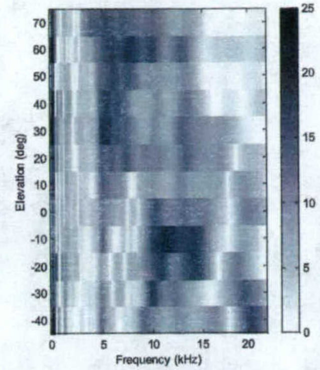
Attenuation (device OFF)



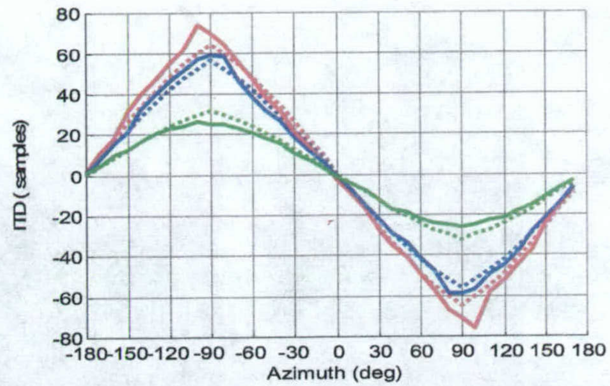
MICH-Sordin frequency response (right ear); all elevations measured at 0° azimuth



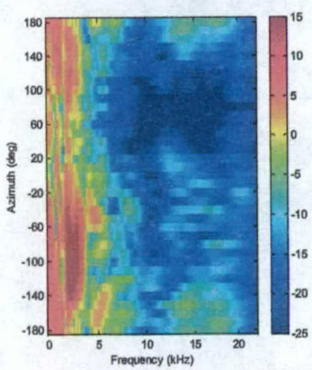
B&K frequency response (right ear); all elevations measured at 0° azimuth



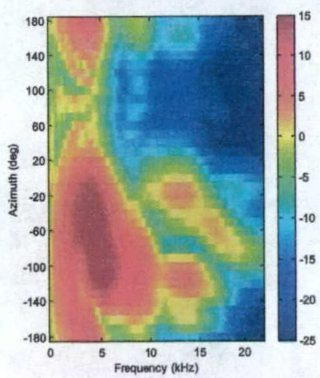
Difference between MICH-Sordin measurements and B&K



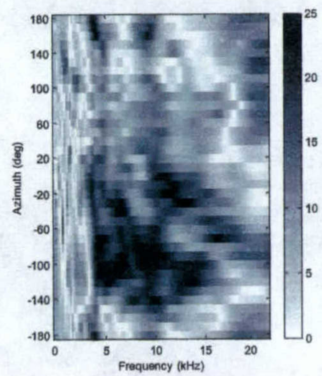
Interaural Time Difference: MICH-Sordin (solid) and B&K (dotted) for elevations +60° (green), 0° (red) and -30° (blue)



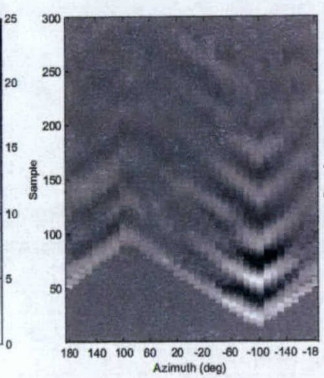
MICH-Sordin frequency response (right ear); all azimuths measured at 0° elevation



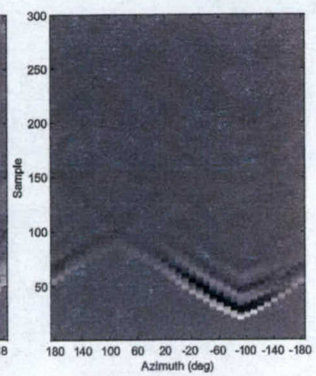
B&K frequency response (right ear); all azimuths measured at 0° elevation



Difference between MICH-Sordin measurements and B&K

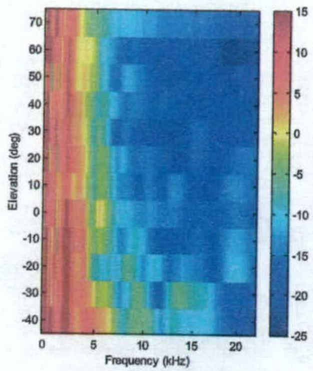


MICH-Sordin time-domain data of 0° elevation ring

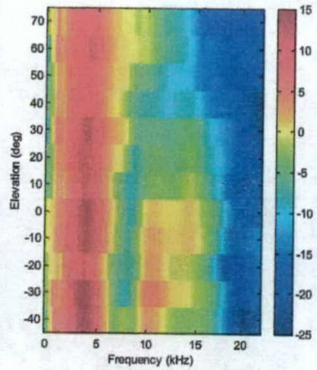


B&K time-domain data of 0° elevation ring

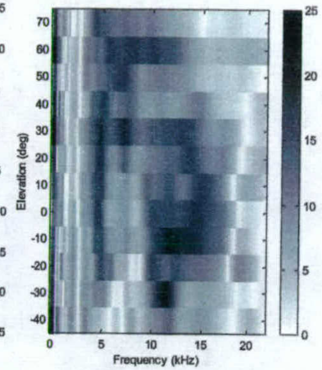
Hear-Thru Active



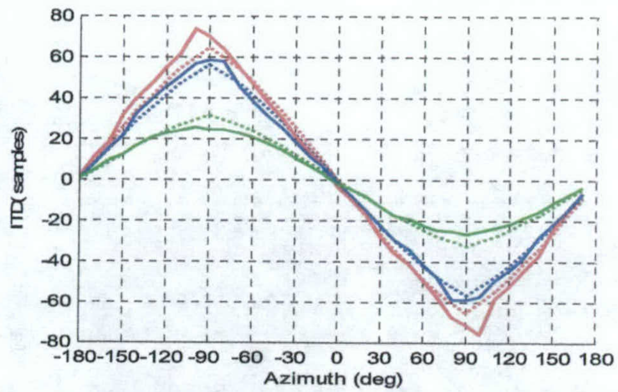
MICH-Sordin frequency response (right ear); all elevations measured at 0° azimuth



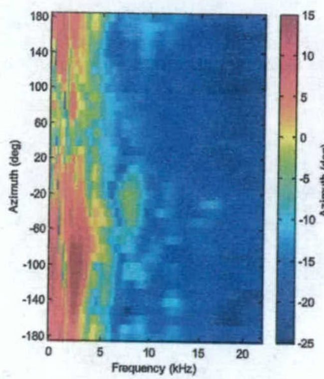
B&K frequency response (right ear); all elevations measured at 0° azimuth



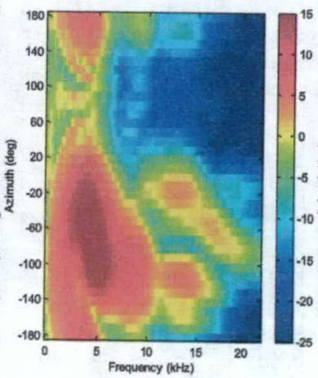
Difference between MICH-Sordin Measurements and B&K



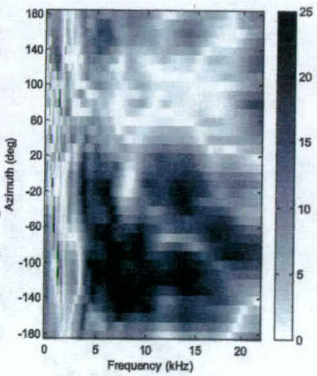
Interaural Time Difference: MICH-Sordin (solid) and B&K (dotted) for elevations +60° (green), 0° (red) and -30° (blue)



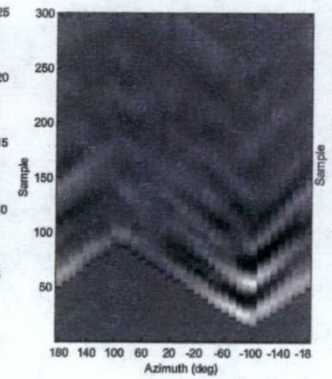
MICH-Sordin frequency response (right ear); all azimuths measured at 0° elevation



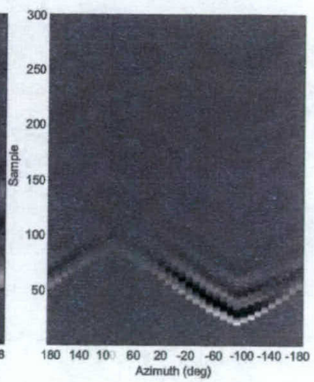
B&K frequency response (right ear); all azimuths measured at 0° elevation



Difference between MICH-Sordin measurements and B&K



MICH-Sordin time-domain data of 0° elevation ring



B&K time-domain data of 0° elevation ring

Appendix E: Subjective COTS Device Assessment

The table on the following pages charts the comments recorded from the subjective tests as described in section 4.3.3.2.

	Comments:	Sound Quality Low	Mid	High	Overall
COTS Active Muffs					
Howard Leight Leightning	Best overall system in terms of comfort, response, natural sound.	Some attenuation. Acceptable	Good	More natural than most other systems	Best
Sordin Supreme III	Something very weird in this system. Center is left and orientation is very unnatural. Not acceptable.	OK	Good	good	phase problem? Unacceptable.
Bilson 707 Impact II	Pretty good in terms of comfort and response, but there seemed to be a balance issue and a difference in high frequency attenuation between the left ear and the right ear. There is distortion in the system at medium voice levels near and far.	Good mid low response, but lacking lower frequency information	~2k bump?	OK	Pretty good
Radians Pro-Amp Electronic Earmuffs	Very good speech recognition, but some unnatural shifts in sound field at apparent threshold of AGC in mid frequencies. Distortion with louder voice levels at close range. No distortion with distant sources.	mostly obscured by extended high end response	seems to be a bump between 2 and 3khz.	lots of information in the 10khz range	Tinny, but good speech recognition
Remington R2000	Very uncomfortable, but the second best in overall response and sound performance. AGC slope seems less steep than others, making them seem more natural sounding.	AGC limits hard below ~200hz but the mid-bass is good. Acceptable	Good	Good	Good
AuSIM Prototypes					
Hard-Pinnae Muffs	Distortion				
Soft-Pinnae Muffs	left mic problem? Different attenuation left to right...				
Passive InEar Plugs					
Foam Plugs	Overall reduction on most bands, and especially in the high frequencies. Muffled.	masked	muffled	muted	OK
Silencio Plugs	Not a lot of reduction in the low end, but natural sounding high end	not much attenuation	OK	OK	
Combat Arms Earplugs	Yellow side has good speech recognition and low end masking. Green side has more of an overall frequency dampening effect.	Yellow - acceptable muted Green - masked	Yellow -good Green - masked	Yellow -good Green - masked	Yellow - Best for speech Green - overall masking

		Comfort	Balance (LR)	Spatial Cues Low Freq	High Freq	Other
COTS Active Muffs						
	Howard Leight Leightning	good	Good	OK	very accurate special orientation	
	Sordin Supreme III	fair	?			No way to rate these except to say they are unacceptable in current configuration.
	Bilson 707 Impact II	good	Poor- different frequency attenuation left to right	Fair	OK	Poor performance based on audible distortion in AGC.
	Radians Pro-Amp Electronic Earmuffs	fair	OK	Obscured by accentuated highs	good	Some clicks and unnatural volume shifts when listener turns in a circle relative to sources.
	Remington R2000	poor - Could not wear these for long.	OK	accurate	good	Too bad these are so uncomfortable. Otherwise a good candidate.
AuSIM Prototypes						
	Hard-Pinnae Muffs					
	Soft-Pinnae Muffs					
Passive InEar Plugs						
	Foam Plugs	good	na	acceptable	OK	
	Silencio Plugs	Fair	na	OK	OK	
	Combat Arms Earplugs	OK	na	OK	Good	

THIS PAGE IS INTENTIONALLY BLANK.

Appendix F: Integration with Dismounted Warfighter Systems

Digital Warfighter

In the 21st century, new personnel systems are being developed and deployed, forever changing capabilities of the modern warfighter. These systems, such as Land Warrior, Objective Force Warrior, FIST, and FIST II, include a network of sensors, data communications and displays, and most importantly digital processing. The audio processing and sensor systems discussed within this document for aural augmentation can integrate tightly within these existing systems with minimal additional cost. Future enhancement of digital electronics will continue to miniaturize these systems and increase energy efficiency.

Audio System

For dismounted soldier applications, the Scorpion risk-reduction program of Natick Soldier Systems identified the following components of the future warfighter audio system to be critical for increased survivability and lethality.

Passive Hearing Protection (muffs and plugs)

Protecting the warfighter's perceptual sensors and orifices from potentially lethal or maiming threats is a special concern for Scorpion. The ears are especially susceptible entry points for biologic, chemical, ballistic, and percussive threats. Additionally, warfighters are exposed to both continuous and impulsive noise at damaging levels as part of normal operations. Passive hearing protection provides a baseline solution to this problem.

Basic Aural Communications Display

To perform basic operations, a warfighter must be able to send and receive aural communications. The baseline communications requirement is the presentation of two concurrent radio signals, typically one in each ear.

Transparent Hearing

Hearing protection and occlusion isolates the warfighter from the environment, deflating situational awareness, confidence, and effectiveness, thus putting the warfighter at high risk and compromising his ability to detect and assess threats. The goal of Scorpion Audio is to at least restore the aural perceptive capability of the soldier such that they can perform tasks equally well with and without hearing protection.

Impulse Noise and Loudness Gating/Compression

Given that transparent hearing provides a controllable sound path circumventing the direct acoustic path to the ear, signals and noises that could possibly damage or impair the warfighter need to be filtered or gated. Potential techniques include limiters, compressors, auto-gates, automatic gain control (AGC), and trims.

Active Noise Reduction (ANR)

Passive noise protection of a small enough size to be worn on a human is physically not as effective at blocking the longer wavelengths of lower frequencies. Low frequency direct path sound must be detected inside the passive protection and have an active canceling signal applied against it, to provide full-spectrum hearing protection.

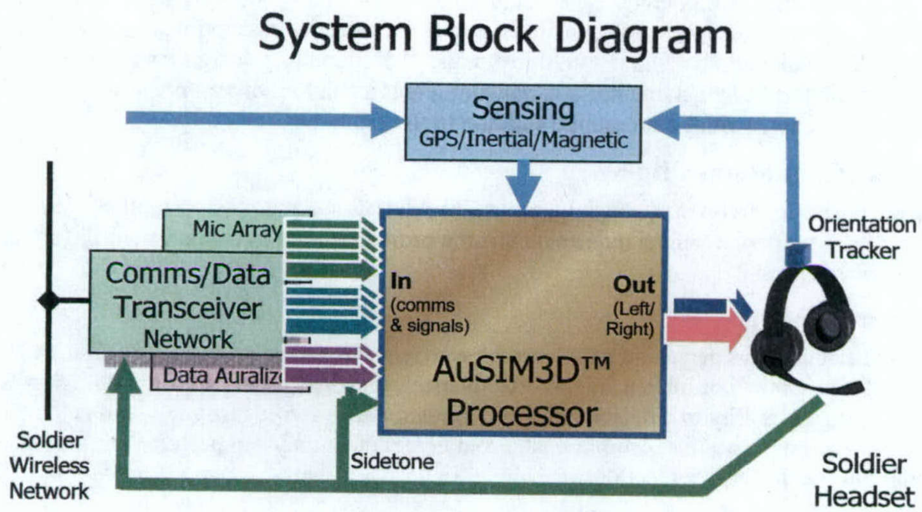
Localized Display of Auralized Information and Data

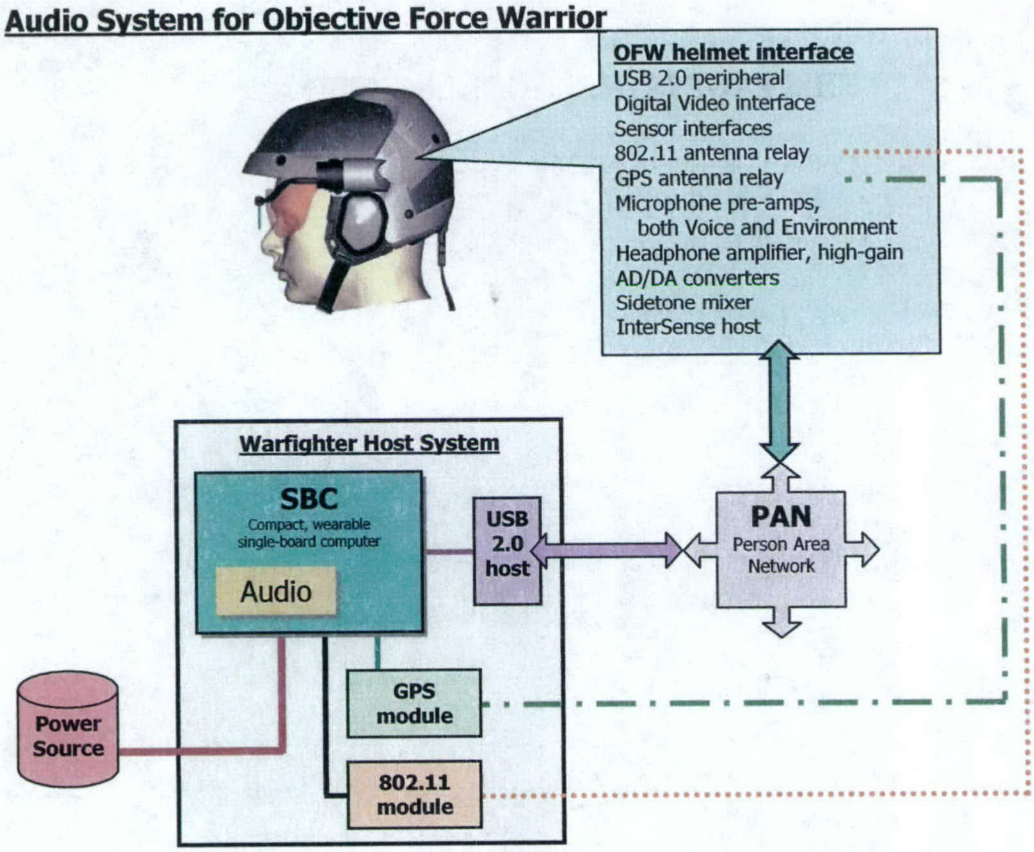
Information cannot be conveniently displayed visually to a dismounted soldier, and, in many circumstances, doing so may compromise their situational awareness (SA). Leveraging aural perception, the warfighter can have potentially large information bandwidth, and remain focused on the task. To keep aural information signals from masking each other, each signal should be spatially independent to provide the human a characteristic for filtering the multiple data streams. Synthetically-generated location cues can be applied to both communication and data auralization. Such displays can leverage orientation tracking and GPS data for spatial coherency and intuitive display of location-inherent data.

Supernormal Listening, including general signal enhancement, selective directional focus, and selective noise suppression

If all of the above objectives are met, then the presentation of the surrounding aural environment is completely controllable and may be specifically augmented with user control. Techniques can provide augmented discrimination of signal from noise, or augmented aural-focusing on a particular direction or signal.

Integration




Audio System for Objective Force Warrior
OFW helmet interface
USB 2.0 peripheral
Digital Video interface
Sensor interfaces
802.11 antenna relay
GPS antenna relay
Microphone pre-amps,
both Voice and Environment
Headphone amplifier, high-gain
AD/DA converters
Sidetone mixer
InterSense host
Warfighter Host System
SBC
Compact, wearable
single-board computer
Audio
USB
2.0
host
PAN
Person Area
Network
GPS
module
802.11
module
Power
Source

THIS PAGE IS INTENTIONALLY BLANK.

9 References

[1] Abel, S.M., Alberti, P.W., Haythornwaite, C., and Riko, K. (1982). "Speech intelligibility in noise: Effects of fluency and hearing protector type," J. Acoust. Soc. Amer. 71:708-715.

[2] Abel, S. M. and Hay, V. H. (1996). "Sound Localization: The Interaction of Aging, Hearing Loss and Hearing Protection," Scand. Audiol. 25(1):3-12.

[3] Abel, S. M., Armstrong, N. M., and Giguere, C. (1993). "Auditory Perception with Level-Dependent Hearing Protectors," Scand. Audiol. 22(2):71-85.

[4] Abel, S. M., Krever, E. M., Giguere, C., and Alberti, P. W. (1991). "Signal Detection and Speech Perception with Level-Dependent Hearing Protectors," J. Otolaryngol. 20(1):46-53.

[5] Agnew, J. (1998). "The causes and effects of distortion and internal noise in hearing aids," Trends in Amplification 3:82-118.

[6] Ahnert, W. (1991). "Binaural Auralization from a Sound System Simulation Program", Preprint of the 91st AES Convention, New York.

[7] Algazi, et al., (2002) "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. 112 (5):2053-2064.

[8] Algazi, R., Duda, D., Thompson, D. M., Avendano, C., (1999) "The CIPIC HRTF Database", Proc. of the 1999 IEEE Workshop on Applications of Signal Processing to Audio Acoustics.

[9] ANSI S3.5-1997, "Methods for Calculation of the Speech Intelligibility Index"

[10] Axelsson, A., Borchgrevink, H., Hamernik, R.P., Hellstrom, P.-A., Henderson, D. and Salvi, R.J. (1996). Scientific Basis of Noise-Induced Hearing Loss, (Thieme, New York).

[11] Barton, G, (1992) "Ambisonics: An Incomplete Glossary", MCS Review, 3(4):10-11, http://members.tripod.com/martin_leese/Ambisonic/reprint007.html

[12] Bauer R.W., Matuzsa J.L., Blackmer R.F., and Glucksberg S. (1966). "Noise localization after unilateral attenuation," J. Acoust. Soc. Am., 40 (2):441-444.

[13] Bell A.J. and Sejnowski T.J. (1995). "An information maximization approach to blind separation and blind deconvolution," Neural Computation 7:1129-1159.

[14] Berger, E.H., Franks, J.R., Lindgren, F. (1996). "International review of field studies of hearing protector attenuation," in *Scientific Basis of Noise-Induced Hearing Loss,* edited by A Axelsson, H. Borchgrevink, R.P. Hamernik, P.-A. Hellstrom, D. Henderson, and R.J. Salvi (Thieme, New York).

[15] Blauert, J, (1997). *Spatial Hearing*, MIT Press.

[16] Bolia, R. S., D'Angelo, W. R., Mishler, P. J., and Morris, L. J. (2001). "Effects of Hearing Protectors on Auditory Localization in Azimuth and Elevation," Human Factors 43:122-128.

[17] Borish, J., J.B. Angell, (1983) "An efficient algorithm for measuring the impulse response using pseudorandom noise", J. Audio Eng. Soc., 31(7):478-488.

[18] Bozzoli, F., Armelloni, E., Ugolotti, E., Farina, A. (2002). "Effects of the background noise on the perceived quality of car audio systems", Preprint of the 112th AES Convention, Munich, Germany.

[19] Bregman, A.S. (1990), Auditory Scene Analysis, MIT Press.

[20] Brungart, D., Rabinowitz, W.M. (1999). "Auditory localization of nearby sources I: Head-related transfer functions," J. Acoust. Soc. Am., 106:1465-1479.

[21] Brungart, D., Durlach, N.I. (1999). "Auditory localization of nearby sources II: Localization of a broadband source in the near field," J. Acoust. Soc. Am., 106:1956-1968.

[22] Brungart, D. (1999). "Auditory localization of nearby sources III: Stimulus," J. Acoust. Soc. Am., 106:3589-3602.

[23] Brungart, D.S., Kordik, A.J., and Simpson, B.D. (2002). "Auditory localization in the horizontal plane with single and double hearing protection," J. Acoust. Soc. Am. 112:2244.

[24] Buell, T.N. and Trahioitis, C. (1997). "Recent experiments concerning the relative potency and interaction of interaural cues," in *Binaural and Spatial Hearing in Real and Virtual Environments,* edited by R.H. Gilkey and T.R.Anderson, (Lawrence Erlbaum Assoc., Mahwah, NJ).

[25] Butler, R.A., Belendiuk, K. (1977). "Spectral cues utilized in the localization of sound in the median saggital plane", J. Acoust. Soc. Am., 61.

[26] Byrne, D. and Noble, W. (1998). "Optimizing sound localization with hearing aids," Trends in Amplification 3:51-73.

[27] Cheshire, S. (1996), "Latency and the Quest for Interactivity". White paper commissioned by Volpe Welty Asset Management, L.L.C., for the Synchronous Person-to-Person Interactive Computing Environments Meeting, San Francisco.

[28] Cotterell, P., (2002) "On the Theory of the Second-Order Soundfield Microphone", Ph.D. Dissertation, University of Reading, http://www.personal.rdg.ac.uk/~shr97psc/Thesis.html

[29] Dancer, A., Grateau, P., Cabanis, A., Barnabe, G., Cagnin, G., Vaillant, T., and Lafont, D. (1992). "Effectiveness of Earplugs in High-Intensity Impulse Noise," J. Acoust. Soc. Am. 91(3):1677-1689.

[30] Daniel, J., (2000) "Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context", Ph.D. Dissertation, http://gyronymo.free.fr/audio3D/download_Thesis_PwPt.html

[31] Daniel, J., "The Experimenter Corner - The Perspective Aspects of Ambisonic and Stereophonic Rendering", http://gyronymo.free.fr/audio3D/download_Thesis_PwPt.html

[32] Davis, R.I., Hamernik, R.P., Ahroon, W.A., and Underwood, K.A. (1996). "Threshold shift dynamics following interrupted impact or continuous noise exposure," in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A Axelsson, H. Borchgrevink, R.P. Hamernik, P.-A. Hellstrom, D. Henderson, and R.J. Salvi (Thieme, New York).

[33] Desloge, J.G, Rabinowitz, W.M. and Zurek, P.M. (1997). "Microphone-array hearing aids with binaural output. I. Fixed-processing systems," IEEE Trans. Speech and Audio Proc. 5:529-542.

[34] Dillon, H. (2001). *Hearing Aids*. Sidney, Australia: Boomerang Press.

[35] Durlach, N.I. (1972). "Binaural Signal Detection: Equalization and Cancellation Theory," in *Foundations of Modern Auditory Theory*, Vol. 2, edited by J.V. Tobias (Academic Press, New York).

[36] Durlach, N.I., Thompson, C.L., and Colburn, H.S. (1981). "Binaural interaction in impaired listeners -- A review of past research," Audiology 20:181-211.

[37] Durlach, N.I. (1991). "Auditory localization in teleoperator and virtual environment systems: ideas, issues, and problems." Perception 20:543-554.

[38] Durlach, N.I. (2003). "Supernormal Listening Systems," accessed January 21, 2003, at http://pellicle.mit.edu/Audio.sls.html)

[39] Durlach, N.I., and Braida, L.D. (1969). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am., 46:372-383.

[40] Durlach, N.I., and Colburn, H.S. (1978). "Binaural phenomena," in *Handbook of Perception*, Vol. 4, edited by E. Carterette and M. Friedman (Academic Press, New York).

[41] Durlach, N.I., Shinn-Cunningham, B.G., and Held, R. (1993). "Supernormal auditory localization. General background," Presence 2:89-103.

[42] Elen, R., (1998) "Ambisonic Surround-Sound in the Age of DVD", Audio Media, http://www.ambisonic.net/ambidvd.html

[43] Evans, M., Angus, J., and Tew, A. (1997) "Spherical harmonic spectra of head-related transfer functions", 103rd AES, preprint 4571.

[44] Evans, M., Tew, A., Agnus, J., (1997) "Spatial Audio Teleconferencing - Which Way is Better?", 4th International Conference on Auditory Displays, http://www.icad.org/websiteV2.0/Conferences/ICAD97/Evans.pdf

[45] Farina, A, (2001). "Re: Tetrahedral to B-format", Sursound Mailing List, http://mail.music.vt.edu/pipermail/sursound/2001-May/007973.html

[46] Farina, A., "Convolution of anechoic music with binaural impulse response", unpublished.

[47] Farina, A., Tronchin, L., (2000). "On the 'Virtual' Reconstruction of Sound Quality of Trumpets," Acoustica, 86: 737-745.

[48] Finchman, L.R., (1985). "Refinements in the impulse testing of loudspeakers", J. Audio Eng. Soc., 33(3):133-140.

[49] Fisher, H., Freedman, S.J. (1968). "The role of the pinnae in auditory localization", Journal of Auditory Research, 8:15-26

[50] Florentine, M. (1976). "Relation between lateralization and loudness in asymmetrical hearing loss," J. Amer. Audiol. Soc. 1:243-251.

[51] Foster, S.H., (1986). "Impulse response measurement using Golay codes", In Proc. ICASSP 86, Tokyo, Japan, 929-932.

[52] Gabriel, K.G., Koehnke, J., and Colburn, H.S. (1992). "Frequency dependence of binaural performance in listeners with impaired binaural hearing," J. Acoust. Soc. Am. 91:336-347.

[53] Gardener, W.G., Martin, K.D. (1995). "HRTF measurements of a KEMAR", J. Acoust. Soc. Am., 97.

[54] Gardner, M.B., Gardner, R.S. (1973). "Problem of localization in the median plane: effect of pinnae cavity occlusion", J. Acoust. Soc. Am., 53.

[55] Gatehouse, S. (1993). "Role of perceptual acclimatization in the selection of frequency responses for hearing aids," J. Amer. Acad. Audiol. 4:296-306.

[56] Golay, M., (1961) "Complementary series", IRE Transfer Info. Theory, 7:82-87.

[57] Greenberg, J.E. and Zurek, P.M. (2001). "Microphone-array hearing aids," in *Microphone Arrays Techniques and Applications*, edited by M.S. Brandstein and D.B. Ward (Springer-Verlag, New York).

[58] Grosveld, (2001) "Binaural Simulation Experiments in the NASA Langley Structural Acoustics Loads and Transmission Facility", NASA/CR-2001-211255, http://techreports.larc.nasa.gov/ltrs/PDF/2001/cr/NASA-2001-cr211255.pdf

[59] Hammershoi, D., Moller, H. (1996). "Sound transmission to and within the human ear", J. Acous. Soc. Am., 100.

[60] Hartmann, W.M. (1997). "Listening in a room and the precedence effect," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R.H. Gilkey and T.R.Anderson, (Lawrence Erlbaum Assoc., Mahwah, NJ).

[61] Hausler, R., Colburn, H.S., and Marr, E. (1983). "Sound localization in subjects with impaired hearing," Acta Oto-Laryngol. Suppl. 400.

[62] Held, R.M. (1955). "Shifts in binaural localization after prolonged exposure to atypical combinations of stimuli," American Journal of Psychology, 68:526-548.

[63] Hofman, P.M., Van Riswick, J.G.A. and Van Opstal, A.J. (1998). "Relearning sound localization with new ears," Nature Neuroscience 1(5): 417-421.

[64] IEC 1998. "Sound system equipment- Part 16: Objective rating of speech intelligibility by speech transmission index". IEC standard 60268-16 second edition 1998.

[65] Johnson, D.H. and Dudgeon, D.E. (1993). *Array Signal Processing* (Prentice-Hall, Englewood Cliffs, New Jersey).

[66] Jot, J.-M., Larcher, V., Pernaux, J.-M., (1999) "A comparative study of 3D audio encoding and rendering techniques", AES 16th Conf. On Spatial Sound Reproduction.

[67] Jot, J.M., Larcher, V., Warusfel, O. (1995). "Digital signal processing issues in the context of binaural and transaural stereophony", Proc. 98th Convention of the Audio Engineering Society, preprint 3980, Paris.

[68] Jot, J.-M., Wardle, S., and Larcher, V. (1998) "Approaches to binaural synthesis", Proc. of the 105th convention of the AES in San Francisco, preprint 4861.

[69] Kassem, S. (1998). "Adapting to auditory localization cues from an enlarged head," M.S. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology: Cambridge, MA.

[70] Katz, B., (1996) "New approach for obtaining individualized head-related transfer functions", J. Acous. Soc. Am., 100(4).

[71] Katz, J. (Ed.) (2002). *Handbook of Clinical Audiology*, 5th Edition. Baltimore, MD: Lippincott Williams & Wilkins.

[72] Keating, D.A., (1996) "The generation of virtual acoustic environments for blind people", Proceedings of the 1st Euro. Conf. Disability, Virtual Reality & Assoc. Tech http://www.google.com/search?q=cache:pHgaIWxLn3UC:www.cyber.rdg.ac.uk/P.Sharkey/WWW/icdvrat/WWW96/Papers/keating.htm+binaural+b-format&hl=en&ie=UTF-8

[73] Kendall, G.S. (1995). "A 3-D Sound Primer: Directional Hearing and Stereo Reproduction", Computer Music Journal, 19(4).

[74] Kirkeby, O., Nelson, P.A., Rubak, P., Farina, A., (1999) "Design of Cross-talk Cancellation Networks by using Fast Deconvolution", 106th AES Convention, Munich, http://pcangelo.eng.unipr.it/Public/Papers/128-AES99.PDF

[75] Kirkup, S.,(1998). The Boundary Element Method in Acoustics, Integrated Sound Software, Todmorden, UK.

[76] Kopco, N., Shinn-Cunningham, B. (2002). "Auditory Localization in Rooms: Acoustic Analysis and Behavior", The 32nd International Acoustical Conference.

[77] Kuhn, G.F., (1987) "Physical acoustics and measurements pertaining to directional hearing", in *Directional Hearing*, Yost W.A. and Gourevitch G.

[78] Lalime, A., (2002) "Development of a Computationally Efficient Binaural Simulation for the Analysis of Structural Acoustic Data", MS Thesis, Virginia Polytechnic Institute and State University http://scholar.lib.vt.edu/theses/available/etd-08142002-122633/unrestricted/Thesis2.pdf

[79] Langendijk, E.H.A. and Bronkhorst, A. (2002). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. 112:1583-1596.

[80] Larcher, V., Jot J.-M., Guyard and Warusfel. (2000). "Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition of HRTF data" 108th AES, preprint 5097.

[81] Letowski, T. R. (1997) "Accuracy of Pointing a Binaural Listening Array," Human Factors Journal, 39(4):651–658.

[82] Letowski, T. R. (1998). "Angular Resolution and Localization with Binaural Microphone Array," 1998 National Conference on Noise Control Engineering, 525–530.

[83] Letowski, T. R. (1997). "Effect of Age and Test Task on Auditory Processing of Time-Compressed Speech," Journal Speech and Hearing Research, 40(5):1192–1200.

[84] Letowski, T. R. (1997). "The Effect on the XM-45 Gas Mask and Hood on Directional Hearing," 41st Annual Meeting of the HFES, 864–867.

[85] Letowski, T. R. (1997). "Perception of Auditory Stimuli in Occupational Safety and Ergonomics," Central Institute for Labor Protection, 1:69–95.

[86] Letowski, T. R. (1998). "Tri-Service Cooperation in Advanced Auditory Display Research," DoD HFE TAG Meeting, 2–3.

[87] Lindley, G. A., Palmer, C. V., Goldstein, H., and Pratt, S. (1997). "Environmental Awareness and Level-Dependent Hearing Protection Devices," Ear and Hearing 18(1):73-82.

[88] Malham, D.G., (1993) "3-D Sound for Virtual Reality Systems Using Ambisonic Techniques", VR93 Conference, http://www2.york.ac.uk/inst/mustech/3d_audio/vr93papr.htm

[89] Malham, D.G., "Ambisonic References," http://www0.york.ac.uk/inst/mustech/3d_audio/ambrefs.htm

[90] Malham, D.G., (1998) "Spatial Hearing Mechanisms and Sound Reproduction" http://www0.york.ac.uk/inst/mustech/3d_audio/ambis2.htm

[91] Markides, A. (1982). *Binaural Hearing Aids*. Academic Press, London.

[92] Mehrgart, S., Mellert, V. (1977). "Transformation characteristics of the external human ear", J. Acoust. Soc. Am., 61.

[93] Mershon, D.H. (1997). "Phenomenal geometry and the measurement of perceived auditory distance," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R.H. Gilkey and T.R.Anderson, (Lawrence Erlbaum Assoc., Mahwah, NJ).

[94] Middlebrooks, J.C. (1999). "Individual Differences in External Ear Transfer Functions Reduced by Scaling in Frequency", J. Acoust. Soc. Am., 106:1480-1492.

[95] Middlebrooks, J.C., Green, D.M. (1991). "Sound localization by human listeners," Annual Review of Psychology, 42:135-159.

[96] Millis, A. W. (1958). "On the minimum audible angle," J. Acoust. Soc. Am., 100:848-856.

[97] Moller, H., Sorensen, M., Hammershoi, D. (1995). "Head-related transfer function of human subjects", J. Audio Eng. Soc., 43(5).

[98] Mooney, J, "Introduction to the Ambisonic Surround Sound System", http://usitweb.shef.ac.uk/~mup01jrm/ambisonics/intro/intro.htm

[99] Moore, B.C.J., (1997). *An Introduction to the Psychology of Hearing*, 4th Edition. San Diego, CA: Academic Press.

[100] Moore, B.C.J., (1998). *Cochlear Hearing Loss*. London: Whurr Publishers Ltd.

[101] Mosko, J. D. and Fletcher, J. L. (1971). "Evaluation of the Gundefender Earplug: Temporary Threshold Shift and Speech Intelligibility," J. Acoust. Soc. Am. 49(6 Part 1):1732-1733.

[102] Mourjopoulos, J.N., (1984). "The removal of reverberation from signals", Ph.D. Dissertation, University of Southampton, UK.

[103] Moy, C., "Technologies for Presentation of Surround-Sound in Headphones", HeadWize Design Series Papers, 1998, 1999, http://headwize2.powerpill.org/tech/sshd_tech.htm

[104] Noble, W. and Byrne, D. (1990). "A comparison of different hearing aid systems for sound localization in the horizontal and vertical planes," Brit. J. Audiol. 24:335-342.

[105] Ohlin, D. (2003). "Cost effectiveness of hearing conservation programs," accessed January 21, 2003 at http://chppm-www.apgea.mil/hcp/costeffective.aspx

[106] Poletti, M, (1997) "Re: Ambisonic encode/decode question", Sursound Mailing List, http://mail.music.vt.edu/pipermail/sursound/1997-March/000012.html

[107] Poletti, M, (1997) "Re: Green's Theorem", Sursound Mailing List, http://mail.music.vt.edu/pipermail/sursound/1997-May/000143.html

[108] Preves, D., Millier, R., Yanz, J., Anderson, B., and Hagen, L. (1998). "A Combination Custom Active Hearing Protector/Hearing Aid," Hearing J. 51(2):34-43.

[109] Rayleigh, Lord [Strutt, J. W.] (1907) "On our perception of sound direction". Philos. Mag. 13:214-232.

[110] Reynosa, M. A., (1999). *The Personnel Armor System Ground Troops Helmet*, Schiffer Publishing Ltd.

[111] Rife, D.D., Vanderkooy, J. (1989). "Transfer function measurement with maximum-length sequence", J. Audio Eng. Soc., 37(6).

[112] Savick, D. S. (1998). "Comparison of Various Types of Head-Related Transfer Functions for 3-D Sound in the Virtual Environment," 66th Military Operations Research Society Symposium (MORSS).

[113] Scanlon, M.V. (1997). "Motion and Sound Monitor and Stimulator," US Patent 5,684,460

[114] Scanlon, M., Tenney, S. (1994), "Binaural Arrays for Hearing Enhancement", J. Acoust. Soc. Am., 96:3262.

[115] Schmidt, R.O. (1986) "Multiple Emitter Location and Signal Parameter Estimation," IEEE Trans. Antennas Propagation, AP-34, 276-280.

[116] Shaw, E. A. G. (1982b). "Hearing Protector Design Concepts and Performance Limitations," in *Personal Hearing Protection in Industry*, edited by P. W. Alberti, Raven Press, New York, NY, 51-68.

[117] Shaw, E.A.G. (1974). "External ear response and sound localization", in *Localization of Sound: Theory and Applications* (Amphora, Groten, CT), 30-41.

[118] Shaw, E.A.G., (1994), "Acoustical features of the external human Head", *in Binaural and spatial hearing in real and virtual environments* , Gilkey R., and Anderson, T. (Eds).

[119] Shinn-Cunningham, B.G. (2000). "Learning Reverberation: Considerations for Spatial Auditory Displays", Proceedings of the 2000 International Conference on Auditory Displays, Atlanta, GA.

[120] Shinn-Cunningham, B.G. (2000). "Adaptation to remapped auditory localization cues; a decision theory model", Perception and Psychophysics, 62(1):33-47.

[121] Shinn-Cunningham, B.G. (2001). "Models of plasticity in spatial auditory processing," Audiology and Neuro-otology, 6(4): 187-191.

[122] Shinn-Cunningham, B.G., Durlach, N.I., and Held, R.M. (1998a). "Adapting to supernormal auditory localization cues. I. Bias and resolution," J. Acoust. Soc. Am., 103:3656-3666.

[123] Shinn-Cunningham, B.G., Durlach, N.I., and Held, R.M. (1998b). "Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response", J. Acoust. Soc. Am., 103:3667-3676.

[124] Shinn-Cunningham, B.G., Streeter, T. and Gyss, J.F. (2001). "Perceptual plasticity in spatial auditory displays," in International Conference on Auditory Display. Espoo, Finland: 181-184.

[125] Shinn-Cunningham, B.G, Lehnert, H., Kramer, G.. Wenzel, E.M., and Durlach, N.I. (1997). "Auditory Displays," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R.H. Gilkey and T.R.Anderson, (Lawrence Erlbaum Assoc., Mahwah, NJ).

[126] Sound-Field Microphones, http://www.soundfield.com

[127] Surround Sound discussion group, http://mail.music.vt.edu/mailman/listinfo/sursound

[128] Steeneken, H.J.M., and Houtgast, T., (1980). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am., 67:318-326.

[129] Streeter, T. and Shinn-Cunningham, B.G. (2002). "A model of short-term perceptual adaptation to remapped spatial auditory cues," in 6th International Conference on Cognitive and Neural Systems, Boston.

[130] Struck, C.J., Temmer, S.F., (1994), "Simulated free field measurements", J. Audio Eng. Soc., 42(6):467-482.

[131] Tokuno, H., Kirkeby, Nelson P.A., Hamada H., (1997) "Inverse Filter of Sound Reproduction Systems Using Regularization", ICIE Trans. Fundamentals, Vol.E80-A, http://search.ieice.or.jp/1997/pdf/a050809.pdf

[132] Travis, C., (1996). "A virtual reality perspective on headphone audio", 101st AES, preprint 4354.

[133] Vanderkooy, J. (1994), "Aspects of MLS measuring systems", J. Audio Eng. Soc., 42(4):219-231.

[134] van Veen, B.D. and Buckley, K.M. (1998). "Beamforming: a versatile approach to spatial filtering," IEEE Acoustics, Speech and Signal Processing Magazine, 5:4-24.

[135] Vause, N.L. and Blank, A.L. (2001). "Creating our own casualties: Auditory effects of anti-tank weapons fire -- A clinical case study," Military Audiology Short Course, Albuquerque, NM.

[136] Vause, N. L. and Grantham, D. W. (1999). "Effects of Earplugs and Protective Headgear on Auditory Localization Ability in the Horizontal Plane," Human Factors 41(2):282-294.

[137] Watson, C.S., (1987). "Uncertainty, informational masking, and the capacity of immediate memory," in *Auditory Processing of Complex Sounds*, edited by W.A. Yost and C.S. Watson (Erlbaum, Hillsdale, NJ).

[138] Wenzel, E.M. (1988b). "Acoustic Origins of Individual Differences in Sound Localization Behavior", J. Acoust. Soc. Amer., 84.

[139] Wenzel, E. M. (1998). "The impact of system latency on dynamic performance in virtual acoustic environments." 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Seattle, WA.

[140] Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L. (1993). "Localization Using Non-individualized Head-Related Transfer Functions," J. Acoust. Soc. Amer., 94.

[141] Wenzel, E., Wightman, F., & Kistler, D. (1991). "Localization with non-individualized virtual acoustic display cues," Proceedings of Human Factors in Computing Systems, CHI '91, pp. 351-359. New York, NY: ACM.

[142] Wightman, F.L., Kistler, D.J. (1989). "Headphones simulation of free-field listening II. Psychophysical validation," J. Acoust. Soc. Amer., 85.

[143] Wightman, F.L., Kistler, D.J. (1992). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Amer. 91:1648-1661.

[144] Yost, W.A. and Nielsen, D.W. (1998). *Fundamentals of Hearing: An Introduction*, 3rd Edition New York: Holt, Rinehart and Winston.

[145] Zahorik, P., (2000), "Limitations in using Golay codes for head-related transfer function measurement", J. Acous. Soc. Amer., 107(3).

[146] Zahorik, P., Tam, C., Wang, K., Bangayan, P., Sundareswaran, V. (2001). "Localization accuracy in 3-D sound displays: the role of visual-feedback training", Proceeding of the Advanced Displays Consortium: ARL's 5th Federated Laboratory Annual Symposium, 17-22.

[147] Zhou, B., Green, D.M., (1992), "Characterization of external ear impulse responses using Golay codes", J. Acous. Soc. Am., 93(2).

[148] Zurek, P.M. (1986). "Consequences of conductive auditory impairment for binaural hearing,"' J. Acoust. Soc. Am., 80:466-472.

[149] Zurek, P.M. (1987). "The precedence effect,'' in *Directional Hearing*, edited by W.A. Yost and G. Gourevitch (Springer-Verlag, New York).

[150] Zurek, P.M. (2000). "A multiple-loudspeaker system for spatial audiometry," presented at Binaural Hearing, Hearing Loss, Hearing Aids, and Cochlear Implants, University of Iowa.

[151] Zurek, P.M., and Delhorne, L.D. (1987). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," J. Acoust. Soc. Am., 82:1548-1559.